

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2002, 22:26:13 ; Search time 22.75 Seconds

(without alignments)
461.936 Million cell updates/sec

Title: US-09-887-038-3

Perfect score: 2426

Sequence: 1 MTWQTLTFAHYQPQOWGHS.....WQPPSKQLPPEAHSDEKM 467

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/Backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	117.5	4.8	3011	3 US-08-811-566-20	Sequence 20, Appl
2	117.5	4.8	3012	3 US-08-811-566-2	Sequence 2, Appl
3	115.5	4.8	505	1 US-08-221-750A-5	Sequence 5, Appl
4	114.5	4.7	2261	4 US-08-444-818-66	Sequence 66, Appl
5	114.5	4.7	2436	4 US-08-444-818-75	Sequence 75, Appl
6	114.5	4.7	2772	4 US-08-444-818-89	Sequence 89, Appl
7	114.5	4.7	2894	2 US-08-466-975A-23	Sequence 23, Appl
8	114.5	4.7	2894	2 US-08-391-671A-23	Sequence 23, Appl
9	114.5	4.7	2894	3 US-08-467-902A-23	Sequence 23, Appl
10	114.5	4.7	2894	4 US-09-275-265-23	Sequence 23, Appl
11	114.5	4.7	2995	4 US-08-444-818-138	Sequence 138, App
12	114.5	4.7	3011	1 US-08-440-103-36	Sequence 36, Appl
13	114.5	4.7	3011	1 US-08-440-542-36	Sequence 36, Appl
14	114.5	4.7	3011	1 US-07-910-760-10	Sequence 10, Appl
15	114.5	4.7	3011	1 US-08-440-519-10	Sequence 10, Appl
16	114.5	4.7	3011	1 US-08-231-368-36	Sequence 36, Appl
17	114.5	4.7	3011	1 US-08-440-210-36	Sequence 36, Appl
18	114.5	4.7	3011	4 US-09-388-874-2	Sequence 2, Appl
19	114.5	4.7	3011	4 US-09-046-604-36	Sequence 36, Appl
20	113.5	4.7	3011	4 US-09-014-416-1	Sequence 1, Appl
21	110.5	4.6	2955	4 US-08-444-818-124	Sequence 124, App
22	110	4.5	3033	1 US-07-925-695-9	Sequence 9, Appl
23	109	4.5	829	4 US-08-444-818-69	Sequence 69, Appl
24	109	4.5	1786	4 US-08-444-818-54	Sequence 54, Appl
25	107.5	4.4	454	4 US-08-444-818-73	Sequence 73, Appl
26	107.5	4.4	663	3 US-08-824-057-3	Sequence 3, Appl
27	107.5	4.4	3011	1 US-08-453-552-2	Sequence 2, Appl

28	107.5	4.4	3011	2 US-08-710-637-2	Sequence 2, Appl
29	107.5	4.4	3011	5 PCT-US93-00907-2	Sequence 2, Appl
30	106.5	4.4	2955	2 US-08-443-260-3	Sequence 3, Appl
31	106.5	4.4	2955	3 US-08-442-805A-3	Sequence 3, Appl
32	106.5	4.4	2955	3 US-08-443-900A-3	Sequence 3, Appl
33	106.5	4.4	2955	4 US-08-249-843-3	Sequence 3, Appl
34	106.5	4.4	3011	2 US-08-833-678A-6	Sequence 6, Appl
35	106.5	4.4	3011	4 US-08-444-818-177	Sequence 177, App
36	106.5	4.4	3011	4 US-08-529-169A-6	Sequence 6, Appl
37	106.5	4.4	3011	5 PCT-US91-02225-10	Sequence 10, Appl
38	106	4.4	3033	1 US-07-925-695-8	Sequence 8, Appl
39	105.5	4.3	3010	4 US-09-014-416-3	Sequence 3, Appl
40	104	4.3	3011	4 US-09-014-416-5	Sequence 5, Appl
41	101.5	4.2	1648	1 US-08-188-281B-12	Sequence 12, Appl
42	101.5	4.2	1648	5 PCT-US94-07280-12	Sequence 12, Appl
43	101.5	4.2	1648	5 PCT-US95-01087-12	Sequence 12, Appl
44	101.5	4.2	3011	1 US-08-188-281B-1	Sequence 1, Appl
45	101.5	4.2	3011	1 US-08-453-552-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-811-566-20
; Sequence 20, Application US/08811566
; Patent No. 6127116
; GENERAL INFORMATION:
; APPLICANT: Rice, Charles et al.
; TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C
; TITLE OF INVENTION: VIRUS (HCV) AND USES THEREOF
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
; STREET: Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/811,566
; FILING DATE: 03-MAR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 1113-1-006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3011 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: N-terminal
; US-08-811-566-20

Query Match 4.8%; Score 117.5; DB 3; Length 3011;
Best Local Similarity 20.2%; Pred. No. 0.022;
Matches 100; Conservative 59; Mismatches 158; Indels 177; Gaps 23;
QY 66 LGIGLAIAAYW--ALISLTIDIDLROATPIHMLVLLYWGVDALATGLSPVRAALVGLA 122

```

Db      705  VGSISASWAIKMEYVVLFLFLDLADARVCSCL-  

QY      123  KLTLVLLVFALARVLNRDRLSLLFSVVITSLFVS  

Db      764  GLVSEFLVFFCFA-----  

QY      183  ADFTSRVSYSLGN-PNLLAALVPTAFS-----  

Db      787  -----VYAFYGMWPLLLLLLALPQRAYALDTEVA  

QY      237  LILTSRGGWLGFAVAMIFWALLGLWFQPRLPAPMR  

Db      831  LSPYKR-----YISNCMMWLQYFLTREVAOLH  

QY      288  -----LIVAVLG-----LEPLRVRLSIF-----  

Db      882  PLIVEDITKLLAIEGPLMLIQASLLKVPYFVRVQGL  

QY      325  LQMIQDRPWLIGIPGNTAFNLVPLYQOARFALSAYS  

Db      939  IK-----LGALTGTYYVNHLLTPLRDMAH-NG  

QY      368  AVEGGLL-----GLTAFAMLLVTAVAVRQVSR-  

Db      991  AACGDIIINGLPVSARRGOETILGPADGMVSKGWRL  

QY      399  RDRN-----PQAFWIMASLAGLAGMLGHGLFDT  

Db      1050 RDKNQVEGEVQIVSTATQTF-LATCINGVCWTV  

QY      446  SFW-QOPPSKOLPP 458  

Db      1109 VGWPAFQGSRSLLTP 1122

```

```

RESULT      2
US-08-811-566-2
: Sequence 2, Application US/08811566
: Patent No. 6127116
:
: GENERAL INFORMATION:
:
: APPLICANT: Rice, Charles et al.
:
: TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C
:
: TITLE OF INVENTION: VIRUS (HCV) AND USES THEREOF
:
: NUMBER OF SEQUENCES: 21
:
: CORRESPONDENCE ADDRESS:
:
: ADDRESSEE: David A. Jackson, Esq.
:
: STREET: 411 Hackensack Ave, Continental Plaza, 4th
:
: STREET: Floor
:
: CITY: Hackensack
:
: STATE: New Jersey
:
: COUNTRY: USA
:
: ZIP: 07601
:
: COMPUTER READABLE FORM:
:
: MEDIUM TYPE: Floppy disk
:
: COMPUTER: IBM PC compatible
:
: OPERATING SYSTEM: PC-DOS/MS-DOS
:
: SOFTWARE: PatentIn Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
:
: APPLICATION NUMBER: US/08/811,566
:
: FILING DATE: 03-MAR-1997
:
: CLASSIFICATION: 435
:
: ATTORNEY/AGENT INFORMATION:
:
: NAME: Jackson Esq., David A.
:
: REGISTRATION NUMBER: 26,742
:
: REFERENCE/DOCKET NUMBER: 1113-1-006
:
: TELECOMMUNICATION INFORMATION:
:
: TELEPHONE: 201-487-5800
:
: TELEFAX: 201-343-1684
:
: INFORMATION FOR SEQ ID NO: 2:
:
: SEQUENCE CHARACTERISTICS:
:
: LENGTH: 3012 amino acids
:
: TYPE: amino acid
:

```

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: N-terminal
US-08-811-566-2

```

Query Match	4.8%;	Score 117.5;	DB 3;	Length 3012;
Best Local Similarity	20.28;	Pred. No. 0.022;		
Matches 100;	Conservative 59;	Mismatches 158;	Indels 177;	Gaps 23;

QY	66	LGGLAIIAAYW--ALLSLTDIDLQOATPIHMLVLLYWGVDALATGLSPVRAAALVGLA	122
		: : : : : : : : : : :	
Db	705	VGSSIASMAIKMEYVLLFLLADARVCSC-LWMMLLSIQAEALLENVILNMAASLAGTH	763
QY	123	KLTLYLLVEALAAARVLNRBRLSLFSVVITSLFVSVYGLNQWITYGVEELATWVDRNSV	182
		↑ : : : : : : : :	
Db	764	GLVSELVFECCA-----WYLGKRWVPGA-----	786
QY	183	ADFTSRVSYLGN-PNIIAAYLPPTAFS-----AAAGVWRGWLPKLIIAATGASSIC	236
		: : : : : : : : :	
Db	787	-----VYAFYGMWPLIIIIIIALLPQRAYALDTEVAASCG-----GVVLVGLMALT	830
QY	237	LITYSRGWLGFWAMLEVMALLGLWFOPLPAPWRWMLFPV-VLGLVAV-----	287
		: : : : : : : : :	
Db	831	LSPYKR-----YISGCMWMLQYFLTRVEAQLHWVWPPLNVRGDAVILLMCVH	881
QY	288	-----LLVAVLG-----LEPLRVRVLSIF-----VGREDDSSNNFRINWMLAV	324
		: : : : : : : : : : :	
Db	882	PLVEDITKLLLAIFGPIMLIQASLLKVPYFVRVQGLRICALARKIAGHY--VQMAI	938
QY	325	LQMIDRPMWLGIGPNTAFNLVPLYQOARFTALSAVSIPLE-----V	367
		: : : : : : : :	
Db	939	IK-----LGALTGTYYVYNHLTPLRDWAH-NGLRDLAAVEEVPVFSRMETKLITWGADT	990
QY	368	AVEGGL-----GLTAFAMLLVTAVTAVRQVSR-----LR	398
		: : : : : : : : : :	
Db	991	AACGDIINGLPVSARRGOEILGPAADMVSKGWRLL-APTAYAOQTRGLLGCTITSLTG	1049
QY	399	RDRN-----PQAFWLMASLAGLGMIGHGLFDIVLYRPEASTL-WMLCIGATA	445
		: : : : : : : :	
Db	1050	RDKNOVEGEVQIVSTATQTF-LATCINGVCWTVYHGAGTRTIASPKGPVIQMTNVDDQL	1108
QY	446	SFW-QPQPSKQLPP	458
		:	
Db	1109	VGWPAFQGSRSRLTP	1122

```

1  RESULT      3
2  US-08-221-750A-5
3  ; Sequence 5, Application US/08221750A
4  ; Patent No. 5643747
5  ;
6  ; GENERAL INFORMATION:
7  ;
8  ; APPLICANT: Baker, Steven M.
9  ; APPLICANT: Deitch, Robert A.
10 ; TITLE OF INVENTION: Genes for the Export of Pertussis
11 ; TITLE OF INVENTION: Holotoxin
12 ; NUMBER OF SEQUENCES: 13
13 ;
14 ; CORRESPONDENCE ADDRESS:
15 ;
16 ; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
17 ;
18 ; STREET: Two Millita Drive
19 ;
20 ; CITY: Lexington
21 ;
22 ; STATE: MA
23 ;
24 ; COUNTRY: USA
25 ;
26 ; ZIP: 02173
27 ;
28 ; COMPUTER READABLE FORM:
29 ;
30 ; MEDIUM TYPE: Floppy disk
31 ;
32 ; COMPUTER: IBM PC compatible
33 ;
34 ; OPERATING SYSTEM: PC-DOS/MS-DOS
35 ;
36 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
37 ;
38 ; CURRENT APPLICATION DATA:
39 ;
40 ; APPLICATION NUMBER: US/08/221,750A
41 ;

```



```

; TITLE OF INVENTION: NANBY Diagnostics and Vaccines
; NUMBER OF SEQUENCES: 777
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,818
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,590
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Ailsa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0110.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508)359-3876
; TELEFAX: (508)359-3885
; INFORMATION FOR SEQ ID NO: 75:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2436 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-444-818-75

```

Query Match 4.7%; Score 114.5; DB 4; Length 2436;
Best Local Similarity 20.6%; Pred. No. 0.033;
Matches 102; Conservative 55; Mismatches 160; Indels 177; Gaps 23;

```

QY 66 LGGLAATAAYW--ALLSTIDIDLRQATPIHMLVLLYWGVDALATGLSPVRAAALVGLA 122
;|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 255 VGSSTASWAIKWEYVLLFLLLADARVCSCSCL-WMMLLSQAELAENLVILMAASLAGTH 313
;|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 123 KLTLVLLVFALAAARVLRNRLRSLFSVVVITSLEFSVYGLNQMIYGVVEELATWVDRNSV 182
;|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 314 GLVSFLVFEFCFA-----WYLGKWPFGA----- 336
;|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 183 ADFTSRVYSYLG-N-PNLLAAYLVPTTAFS-----AAAGVWRGWLPKLLAIAATGASSLC 236
;|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 337 -----VYTFYGMWPLLLLLLPQRAYALDTEVAASCG-----GVVLGIMALT 380
;|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 237 LILTYSRGGMGFMAMIFVWALLGLYWFQPRLPAPWRRLFPV-VLGLVAV----- 287
;|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 381 LSPYYKR-----YISWCLMWLQYFLTRVEAQLHWIIPPLNVGRGDAVILMCAYH 431
;|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 288 -----LLVAVLG-----LEPLRVVLSIF---VGREDDSSNNFRINWTLAV 324
;|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 432 PFLVFDITKLLAVFGPLWILQASLKLKVPYFVRVQGLRFCALARKMIGGHY----- 483
;|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 325 LQMIQDRPWLTGIGPGNTAFNLVPLVQARFTALSAVSPL-----V 367
;|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 484 VQMTIIR--LGAALTGYVYVNHLPRLDMAH-NGLRDLAVAVEPVEFSQMETKLITWGADT 540
;|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 368 AVEGGL-----GLTAFAMLLLVTAVTAVROVSR-----LR 398
;|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 541 AACGDIINGLPSARARGREILLGPADGWSKWRLL-APITAYAOOTRGLGCIITSLTG 599
;|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 399 RDRN-----PQAFWLMASLAGLMLGHLFDTVLYRPEASTL-WMLCIGATA 445
;|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 600 RDKNOVEGEOIVSTAQTF-LATCINGVCWTVYHGAGTRTIASRGKVPYIQMYTNVDQL 658
;|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

QY 446 SFW-QPQPSKQLPP 458
DB 659 VQWPAPQGSRSLLTP 672

```

RESULT 6
US-08-444-818-89
; Sequence 89, Application US/08444818
; Patent No. 6150087
; GENERAL INFORMATION:
; APPLICANT: Chien, David Y.
; APPLICANT: Rutter, William J.
; TITLE OF INVENTION: NANBY Diagnostics and Vaccines
; NUMBER OF SEQUENCES: 777
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,818
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,590
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Ailsa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0110.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508)359-3876
; TELEFAX: (508)359-3885
; INFORMATION FOR SEQ ID NO: 89:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2772 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-444-818-89

```

Query Match 4.7%; Score 114.5; DB 4; Length 2772;
Best Local Similarity 20.6%; Pred. No. 0.04;
Matches 102; Conservative 55; Mismatches 160; Indels 177; Gaps 23;

```

QY 66 LGGLAATAAYW--ALLSTIDIDLRQATPIHMLVLLYWGVDALATGLSPVRAAALVGLA 122
;|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 591 VGSSTASWAIKWEYVLLFLLLADARVCSCSCL-WMMLLSQAELAENLVILMAASLAGTH 649
;|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 123 KLTLVLLVFALAAARVLRNRLRSLFSVVVITSLEFSVYGLNQMIYGVVEELATWVDRNSV 182
;|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 650 GLVSFLVFEFCFA-----WYLGKWPFGA----- 672
;|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 183 ADFTSRVYSYLG-N-PNLLAAYLVPTTAFS-----AAAGVWRGWLPKLLAIAATGASSLC 236
;|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 673 -----VYTFYGMWPLLLLLLPQRAYALDTEVAASCG-----GVVLGIMALT 716
;|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 237 LILTYSRGGMGFMAMIFVWALLGLYWFQPRLPAPWRRLFPV-VLGLVAV----- 287
;|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 717 LSPYYKR-----YISWCLMWLQYFLTRVEAQLHWIIPPLNVGRGDAVILMCAYH 767
;|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 288 -----LLVAVLG-----LEPLRVVLSIF---VGREDDSSNNFRINWTLAV 324
;|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 768 PFLVFDITKLLAVFGPLWILQASLKLKVPYFVRVQGLRFCALARKMIGGHY----- 819
;|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```



```

OY      325 LQMIODRPLWGLGPGNTAFNLVPLYOOARFTALSAYSVPLE-----V    367
          :|||: || | : || | : || | : || | : || | : || | : || |
Db     820 VQMVIK--LGALTGTYYVNHLTLPRLDMAN-NGLRDLAVAAVEPYVSOMETKLITWGADT   876
OY      368 AVEGGL-----GITAFAWLLLVTAIVRAVOYSR-----LR    398
          | | :: | : | : | : | : | : | : | : | : | : |
Db     877 AACGDIIINGLPSARRGREILLGPADGVSKGMWRL-APITYAQQOTRGILCILTSLTG   935
OY      399 RDRN-----POAFWLMAISLAGMLGHGLEFDVTLYVRPEASTJ-WWLICIGAI A    445
          ||:: | | | : | : | : | : | : | : | : | : | : |
Db     936 RDKNQEVEGOIYSTAQTF-LATFNCINVCWTYVHGAGTRFIASPKPVIOMTYTNVDQL   994
OY      446 SFW-QOPSRKOLPP 458
          | || | : | | 
Db     995 VGWPAPQGSRSLTP 1008

RESULT       7
US-08-466-975A-23
; Sequence 23, Application US/08466975A
; Patent No. 5910404
; GENERAL INFORMATION:
; APPLICANT: DELEYS, ROBERT J
; APPLICANT: POLLET, DIRK
; APPLICANT: MAERTENS, GEERT
; APPLICANT: VAN HEUVERSMUN, HUGO
; TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
;                   TITLES OF INVENTIONS TO HEPATITIS C VIRUS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESSES:
; ADDRESSSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GIEBE ROAD
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,975A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/391,671
; FILING DATE:
; APPLICATION NUMBER: US 07/920,286
; FILING DATE: 14-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/EP91/02409
; FILING DATE: 13-DEC-1991
; APPLICATION NUMBER: EP 90124241.2
; FILING DATE: 14-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: SADOFF, B.J.
; REGISTRATION NUMBER: 36,663
; REFERENCE/DOCKET NUMBER: 1487-5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 7038164000
; TELEFAX: 7038164100
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2894 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
US-08-466-975A-23
```

```

Query Match          4.7%; Score 114.5; DB 2; Length 2894;
Best Local Similarity 20.6%; Pred. No. 0.042;
Matches 102; Conservative 55; Mismatches 160; Indels 177; Gaps 23;

QY      66 LGLGLAIAAYW---ALLSLTDIDIRQATPIHMLVLLYGVDALATGLSPVRAALVGLA 122
       :|::||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      705 VGSISASWAIKWEYVLLFLLLADARVCSC-LMMLLISOAEALENLVIILNASSLAGTH 763
QY      123 KLTLYLLEFALAARVLARNPRLSRLLFSVVITSLFVSYYGLNQMIYGVEELATWDRNSV 182
       |::||| | | | | | | | | | | | | | | | | | | | | | | | |
Db      764 GLVSFLVFECFA-----WYLKKGWVPGA-----786
QY      183 ADFTSRVYSYIGN-PNLAAAYLVPTTAFS----AAAGVWRGWLPKLLAIATAAGASSLC 236
       ||::||| | | | | | | | | | | | | | | | | | | | | | |
Db      787 -----VYTFGMWPPLLLLLLALPORAVALDTEVAASC-----GVVLVGLMALT 830
QY      237 LIITYSRGGWLGFVAMIEFWALLGLYWFQPRLPAPWRRLFPV-VLGGLVAV-----287
       | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      831 LSPYKR-----YISWCLMWLOYFLTREAOHLHWIPLNVGRDAVILLMCAPH 881
QY      288 -----LIVAVLG-----LEPLRVRVLSIF-----VGEDSSNNFRINWLA 324
       ||:||| | | | | | | | | | | | | | | | | | | | | |
Db      882 PTLVFDITKLLEAVEGPLMILDASLKVPYFVRVQGLBPCALARKMIGHY-----933
QY      325 LQMIDRPWLIGPGNTAFNLVYPYQAARTALSAYSVDL-----V 367
       ::||: | | | | | | | | | | | | | | | | | | | | |
Db      934 VQMVLIK--LGALTGTYYVNHLFPLERDMAH-NGLRDLAAVEPEVFSOMETKLITWGADT 990
QY      368 AVEGGL-----GLTAFAWLLVYAVTAVRQVRS-----LR 398
       | | | | | | | | | | | | | | | | | | | | | |
Db      991 AACGDITNGLEVSARRGREIILGPADGMVSKGWRLL-APITAYAQT RGLIGCTITSLTG 1049
QY      399 RDNRN-----PQAFWLMSLAGLGMIGHGLFDTVLYRPEASTL-WMLCIGAIA 445
       ||::| | | | | | | | | | | | | | | | | | | | |
Db      1050 RDKNOVEGEVOIVSTAOTF-LATCINGVCWTYVHGAGRRTIASPKGPVIQMTNVDDOL 1108
QY      446 SFW-QPQPSKQLPP 458
       | | | | | | | | | | | | | | | | | | | | |
Db      1109 VGWPAPQGSRSULTP 1122

RESULT      8
US-08-391-671A-23
; Sequence 23, Application US/08391671A
; Patent No. 5922532
; GENERAL INFORMATION:
; APPLICANT: DELEYS, ROBERT J
; APPLICANT: POLLET, DIRK
; APPLICANT: MAERTENS, GEERT
; APPLICANT: VAN HEUVERSWON, HUGO
; TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
; TITLE OF INVENTION: ANTIBODIES TO HEPATITIS C VIRUS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/391,671A
; FILING DATE: 21-FEB-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/920,286

```

; FILING DATE: 14-OCT-1992
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: WO PCT/EP91/02409
 ; FILING DATE: 13-DEC-1991
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: EP 90124241.2
 ; FILING DATE: 14-DEC-1990
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: SADOFF, B.J.
 ; REGISTRATION NUMBER: 36,663
 ; REFERENCE/DOCKET NUMBER: 1487-5
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 7038164000
 ; TELEFAX: 7038164100
 ; INFORMATION FOR SEQ ID NO: 23:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2894 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ;

Query Match	4.7%;	Score 114.5;	DB 2;	Length 2894;
Best Local Similarity	20.68;	Pred. No. 0.042;		
Matches 102; Conservative	55;	Mismatches 160;	Indels 177;	Gaps 23;

```

0Y . 66 LGLGAAIAAYW---ALLSTLDDLRQATPIHMLVLLLYMGVDALATGLSPVRAALVGLA 122
    :|:|:| | | | | | | :|:|:| | | | | |
Db 705 VGSISASWAIKWEYVVLFLLELDADARVCSC-LMMALLISOEALENILVILNAAIAGTH 763
OY 123 KLTLYLVEFALAARVLRMNPRLRSLLFSVVVITSIFSVSYGLNMÖIYGVEELATWDNRNV 182
    -|:|:| | | | | | | | | | | | | |
Db 764 GLVSFLVEFCFA-----WYLKKGWVPGA----- 786
OY 183 ADFTSRVSYLGN-PNLAAVLPPTAFS----AAATGWRGWLPKLDAIATGASSLC 236
    ||:| | | | | | | | | | | | | |
Db 787 -----VFTEFGMWPLLLLLLALPÖRAYALDFEVAASC-----GVVLVGMLALT 830
OY 237 LILTYSRCGMIGFVAMIEFNALLGLYWFQPRLPAPWRKWLFPV-VLGGLVAV----- 287
    | | | | | | | | | | | | | | | |
Db 831 LSPYYKR-----YISCMWMLOFYLTRAEQLHWMIPLPNVRGRDAVILLMCAYH 881
OY 288 -----LLVAVLG-----LEPLRVRLSIF---VGREDSNNFRINWLVAV 324
    ||:| | | | | | | | | | | | | |
Db 882 PTLVFEDITKLLLAVEGPLMILDASLLKVPEFVRQGLLRFICALARKMIGHY----- 933
OY 325 LQMIDRPMWIGIPGNTAFNLVYPPLYQÖARFTALSAYSIPLE-----V 367
    :||:| | | | | | | | | | | | | |
Db 934 VÖMVIIR- LGALTGTYYVYNHLTPLRDWAH-NGLRDLAVAVEPVFSÖMETKLTWGADT 990
OY 368 AVEGGLL-----GLTAFAWLLLVATAVAVRÖYSR-----LR 398
    -| | | | | | | | | | | | | |
Db 991 AACGDITNCLPVSARRGREILLGPADGMVSKGWRLL-APITAYAÖQTRGLGCIIITSLTG 1049
OY 399 RDRN-----PÖAFLMASLAGLGMIGHGLFDTVLYRPESATLL-WWLCIGAIA 445
    ||:| | | | | | | | | | | | | |
Db 1050 RDKNÖVEGEVGIYSTAQTF-LATCINGVCWTYHGAGTRFIASPKGPVIÖMTNWDDLL 1108
OY 446 SFW-QÖPPSKÖLPP 458
    | | | | | | | |
Db 1109 VGMPAPOGSRSLTP 1122
```

RESULT 9
US-08-467-902A-23
; Sequence 23, Application US/08467902A
; Patent No. 6007982
; GENERAL INFORMATION:
; APPLICANT: DELEYS, ROBERT J

APPLICANT: POLLET, DIRK
APPLICANT: MAERTENS, GEERT
APPLICANT: VAN HEUVERSWUN, HUGO
TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
TITLE OF INVENTION: ANTIBODIES TO HEPATITIS C VIRUS
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESSES:
ADDRESSEE: NIXON & VANDERHYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22201
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/467,902A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/391,671
FILING DATE:
APPLICATION NUMBER: US 07/920,286
FILING DATE: 14-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/EP91/02409
FILING DATE: 13-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 90124241.2
FILING DATE: 14-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B.J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 1487-5
TELECOMMUNICATION INFORMATION:
TELEPHONE: 7038164000
TELEFAX: 7038164100
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 2894 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-467-902A-23

Query Match	4.7%;	Score 114.5;	DB 3;	Length 2894;
Best Local Similarity	20.68;	Pred. No. 0.042;		
Matches 102;	Conservative 55;	Mismatches 160;	Indels 177;	Gaps 23;

```

QY      66 LGLGLAIAIAW---ALLSLTDIDLEKQATPIHWLVLLYWCVAALATGLSPVRAALAVGLA 122
      : : : : : | | | : : : : : | | | : : : : :
Db      705 VGSISASWAIKWEYVVLFLLLADARVCSC-L-MMLLISQAEALLENLVIILMAASLAGTH 763
QY      123 KLLVLLVPAALAAVRLRNPRLRSLSFSVVVITSLFVSVYGLNOMIYGVBELATWVDRNSV 182
      | : : : | | | | | | | | | | | | | | |
Db      764 GLVSFLVPECPA-----WYLKQKWVPGA----- 786
QY      183 ADFTSRVYSYIGN-PNLLAAVLYPTAFS-----AAAGVRCGMLPKLLAIAATGASSLC 236
      || : : | | | | : : : : | | : : |
Db      787 -----VYTFEGMWPLLLLLLALPQRAYALDTEVAASCG-----GVVLVGLMALT 830
QY      237 LILYTRSGCWLGFVAMIFWVALLGLYWFQRLPAPWRRWLFVY-VLGGLVAV----- 287
      | | | | | | | | : : | : : | : : | : : |
Db      831 LSPYKRR-----YISWCLMWLOYFLTRVEAQHLHWIPEPLNVRGRDAVILLMCVH 881
QY      288 -----LLVAVLG-----LEPLRVRLSIF-----VGREDDSSNFRINWMLAV 324

```

Db 882 PTLVFDITKLLAVGCPMLILDASLKVPRFVRVQGLLRFCALARKMIGHY----- 933
QY 325 LQMIODRPMWLGIGPGNTAFNLVYPLYQOARFTALSAVSPL-----V 367
Db 934 VQMWIK--LGALTGTYYVNHLPRLDMAH-NGLRDLAVAVEPVPVSOMETKLITWGADT 990
QY 368 AVEGGL-----GLTAFAMLLVTAVTAVROVSR-----LR 398
Db 991 AACGDIINGLPVSARRGREILGPADGMVSKGWRLL-APITAYAQOTRGLLCITSLTG 1049
QY 399 RDRN-----POAFWLMASLAGMLGHLEFDTVLYRPEASTL-WMLCIGAIA 445
Db 1050 RDKNQEVEGVQIVSTAQTF-LATCINGVCWTVYHGAGTRTIASPKGPVIQMTNVQDL 1108
QY 446 SFW-QPQPSKQLPP 458
Db 1109 VGWPAPOGSRSLTP 1122

RESULT 10

US-09-275-265-23

; Sequence 23, Application US/09275265
; Patent No. 6287761

GENERAL INFORMATION:

APPLICANT: DELEYS, ROBERT J

APPLICANT: POLLET, DIRK

APPLICANT: MAERTENS, GEERT

APPLICANT: VAN HEUVERSWUN, HUGO

TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF

NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHYE P.C.

STREET: 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VA

COUNTRY: USA

ZIP: 22201

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/275,265

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/391,671

FILING DATE: 21-FEB-1995

APPLICATION NUMBER: US 07/920,286

FILING DATE: 14-OCT-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO PCT/EP91/02409

FILING DATE: 13-DEC-1991

APPLICATION NUMBER: EP 90124241.2

FILING DATE: 14-DEC-1990

ATTORNEY/AGENT INFORMATION:

NAME: SADOFF, B.J.

REGISTRATION NUMBER: 36,663

REFERENCE/DOCKET NUMBER: 1487-5

TELECOMMUNICATION INFORMATION:

TELEPHONE: 7038164000

TELEFAX: 7038164100

INFORMATION FOR SEQ ID NO: 23:

SEQUENCE CHARACTERISTICS:

LENGTH: 2894 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

HYPOTHETICAL: NO
ANTI-SENSE: NO
US-09-275-265-23

Query Match 4.7%; Score 114.5; DB 4; length 2894;
Best Local Similarity 20.6%; Pred. No. 0.042;
Matches 102; Conservative 55; Mismatches 160; Indels 177; Gaps 23;

QY 66 LGIGLAIAIAYW---ALLSTDIDLQATPIHMLVLVYGVDA LATGLSPVRAALVGLA 122
Db 705 VGSSTASWAIKWEYVLLFLLLADARVCSC-LWMLLISQAEALLENVILNAA SLAGTH 763
QY 123 KLTLYLVEFALARVLRNPRLSLFSVVVITSLFVSYYGLNQMIYVEELATVDRNSV 182
Db 764 GLVSTLVFFCPA-----WYLGKQWVPGA----- 786
QY 183 ADFTSRVYSYLG-NPMLAAYVPTTAPS-----AAAGVWRGWLPKLAIATGASSLC 236
Db 787 -----VTFYGMWPLLLALLALFORAYALDTEVAASCG-----GVVLVGLMALIT 830
QY 237 LILTYSRGCGWLGFMAMIFWALLGLYWFOPRLPAPRRWLPV-VLGLVAV----- 287
Db 831 LSPYKR-----YISWCLMWLYFELTRVEAQLHWIPLNVRGGRDAVILMCVH 881
QY 288 -----LLVAVLG-----LEPLRVVLSIF-----VGREDDSSNFRINWILAV 324
Db 882 PTLVFDITKLLAVGCPMLILDASLKVPRFVRVQGLLRFCALARKMIGHY----- 933
QY 325 LQMIODRPMWLGIGPGNTAFNLVYPLYQOARFTALSAVSPL-----V 367
Db 934 VQMWIK--LGALTGTYYVNHLPRLDMAH-NGLRDLAVAVEPVPVSOMETKLITWGADT 990
QY 368 AVEGGL-----GLTAFAMLLVTAVTAVROVSR-----LR 398
Db 991 AACGDIINGLPVSARRGREILGPADGMVSKGWRLL-APITAYAQOTRGLLCITSLTG 1049
QY 399 RDRN-----POAFWLMASLAGMLGHLEFDTVLYRPEASTL-WMLCIGAIA 445
Db 1050 RDKNQEVEGVQIVSTAQTF-LATCINGVCWTVYHGAGTRTIASPKGPVIQMTNVQDL 1108
QY 446 SFW-QPQPSKQLPP 458
Db 1109 VGWPAPOGSRSLTP 1122

RESULT 11

US-08-444-818-138

; Sequence 138, Application US/08444818
; Patent No. 6150087

GENERAL INFORMATION:

APPLICANT: Chien, David Y.

APPLICANT: Rutter, William J.

TITLE OF INVENTION: NANBV Diagnostics and vaccines

NUMBER OF SEQUENCES: 777

CORRESPONDENCE ADDRESS:

ADDRESSEE: Chiron Corporation

STREET: 4560 Horton Street

CITY: Emeryville

STATE: CA

COUNTRY: USA

ZIP: 94608-2916

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/444,818

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/403,590


```
QY      237 LITFSRGGWLGFEVAMIFVALLGLYWFQPRLPAPWRRLPEPV-VLGLVAV----- 287
          |   |   |   |   |   |   |   |   |   |   |   |   |
Db      831 LSPYYKR-----YISWCLMWLOFLTRVEAQLHWIPLNVRGGRDAVILIMCAVH 881
QY      268 -----LLVAVLG-----LEPLRVRLSIF---VGREDSSNFERINWLAV 324
          ||:|||   |   |   |   |   |   |   |   |   |   |
Db      882 PLVEDITKLLAVEGPLMILOASLLKVPYFVRVOGLRECALARKMIGHY----- 933
QY      325 LQMIDRPMLIGIPGNATFNLVYPLOYOARFTALSAVSYLE-----V 367
          :||:  :  ||   |   |   |   |   |   |   |   |   |
Db      934 VQMWLIK-LGALTGTYYVNHLTPLRDMAH-NGLRDLAAVEPVPVESOMETKLITWGADI 990
QY      368 AVEGGLL-----GLTAFAWLLLVTAVTAVRQVSR-----LR 398
          |   |   |   |   |   |   |   |   |   |   |   |
Db      991 AACGDIIINGLPVSARRGREILLGPADGMVSKGWRL-APTAYAOQTROLLGCITSLTG 1049
QY      399 KDRN-----PQAFWLMAFLAGLAGMLGHGLEFDIVLYLRPEASTL-WWLCICATA 445
          ||:||   |   |   |   |   |   |   |   |   |   |
Db     1050 RDKNOVEGEVQIVSTAQTF-LATCINGVCMTVIYHGAGIRTIASPKGPVIOMYTNVDOL 1108
QY      446 SFW-QPQPSKOPLP 458
          |   |||   |   |   |   |   |   |   |   |   |
Db     1109 VGMPAPQGSRSLTP 1122
```

RESULT 15
US-08-440-519-10
; Sequence 10, Application US/08440519
; Patent No. 5712087

GENERAL INFORMATION:

APPLICANT: Houghton, Michael
 APPLICANT: Choo, Qui-Lim
 APPLICANT: Kuo, George
 TITLE OF INVENTION: Combinations of Hepatitis C virus (HCV)
 TITLE OF INVENTION: Antigens for Use in Immunoassays for Anti-HCV Antibodies
 NUMBER OF SEQUENCES: 12
 CORRESPONDENCE ADDRESS:

ADDRESSEE: Chiron Corporation
STREET: P.O. Box 8097 (Int. Prop. R-440)

```

;
; CITY: Emeryville
; STATE: CA

```

COUNTRY: U.S.A.
ZIP: 94662-8097

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

```

```

; SOFWMANE.  FALCULIN RELEASE #1.0;
CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: US/08/440,519

FILING DATE: 12-MAY-19

CLASSIFICATION: 530

PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 02 0310 7600

APPLICATION NUMBER: US 0
FILING DATE: 07-MAR-1993

ATTORNEY/AGENT INFORMATION:
FILING DATE: 07-JUL-1992

NAME: Blackburn Esq., Robert P.
;
; AIDORNEI/ASINI INFORMATION;
;

REGISTRATION NUMBER: 30,447

REFERENCE/DOCKET NUMBER: 0101.0022 ;

TELECOMMUNICATION INFORMATION ;

TELEPHONE: (510) 601-2702

TELEFAX: (510) 655-3542

```

; INFORMATION FOR SEQ ID NO: 10:
;
;     SEQUENCE CHARACTERISTICS:
;

```

SEQUENCE CHARACTERISTICS:

LENGTH: 3011 AMINO ACIDS
TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-440-519-10

Query Match 4.78; Score 114.5; DB 1; Length 3011;

Best Local Similarity 20.6%; Pred. No. 0.045;
Matches 102; Conservative 55; Mismatches 160; Indels 177; Gaps 23;

```

QY 66 LGLGIATAAYW---ALISLTDIDLQOATPIHMLVLLYWGVDALATGLSPVRALVGLA 122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 705 VGSSTASWAIKWEYVLLFLLADARVCSC-LMMLLISQAEALENLVILNASTLATH 763

QY 123 KLTLYLVFALAARVLERNRSLFSVVVITSLEVSYYGLNOMITYGVEELATVDRNSV 182
   | : : | | | | | | | | | | | | | | | | | | | | | |
Db 764 GLVSFLVFECFA-----WYLGKMWVPGA----- 786

QY 183 ADFTSRVYSYLG-N-PNLLAAYLVPTTAFS-----AAAIGWRGWLPKLLAIAATGASSLC 236
   || : : | | | | : | : : | : : | : : | : : | : : |
Db 787 -----VYTFYGMWPLILLLLALPORAVALDTEVAASCG-----GVVLGIMALT 830

QY 237 LILTYSRGWLGFVAMIFWALLGLYWFOPRLPAPWRBMLFPV-VLGLVAV----- 287
   | | | | | | | | | | | | | | | | | | | | | |
Db 831 LSPYKR-----YISMCMLWLQYFLTRVEAQLHWIPLNVRGRDAVILLMCAVH 881

QY 288 -----LLVAVLG-----LEPLRVRLSIF-----VGREDDSSNFRINWLA 324
   : : : : : : : : : : : : : : : : : : : : : :
Db 882 PLLVEDITKLLLAVFGLMLLQASLLKVPFYFVRVQGLRFCALARKMIGHY----- 933

QY 325 LOMIQDRPWLGTGPGNTAFNLVYPLYQOARETALSAYSPLE-----V 367
   : || : : | | : : : : | | : : : : | : : : : :
Db 934 VQMWIIR-LGALTGTYYVNHLLPLRDMAH-NGLRDLAVALVEPVVFSQMETKLTWGADT 990

QY 368 AVEGGL-----GLTAFAMLLVTAVTAVROVSR-----LR 398
   | | : : | : : | : : | : : | : : | : : | : :
Db 991 AACGDIIINGLPVSARRRGRETIILGPADGMVSKGWRL-APITAYAOOTRGLIGCITISLTG 1049

QY 399 RDRN-----PQAFWLMASLAGLGMIGHLEDTVLVYRPEASTL-WMLCIGATA 445
   || : : | | | | : : : : : : : : : : : : : :
Db 1050 RDKNOVEGEVQIVSTAQTF-LATCINGVCWTVYHGAGTRTIASPKGPVIQMYTNVDL 1108

QY 446 SFW-QPQPSKOLPP 458
   | | | | : : | |
Db 1109 VGMFAPQGSRSRSLTP 1122

```

Search completed: January 10, 2002, 23:04:50
Job time: 2317 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2002, 22:25:53 ; Search time 38.09 Seconds
(without alignments)
908.171 Million cell updates/sec

Title: US-09-887-038-3
Perfect score: 2426
Sequence: 1 MTVMQTLTFPAHYQPQMGHS.....WQPPQPSKQLPPEAEHSDEK 467

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_1101:*
1: /SIDS2/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	146	6.0	388	20	AAV22569 Bacterial general
2	143.5	5.9	348	21	AAV81520 Streptococcus pneu
3	132.5	5.5	758	22	AAB79584 Corynebacterium gl
4	132.5	5.5	962	22	AAAG0046 C glutamicum prote
5	130	5.4	536	21	AAAB4555 Virulence gene pro
6	127	5.2	461	18	AAW19276 pseudomonas aerugi
7	124.5	5.1	483	22	AAAG8934 C glutamicum prote
8	124.5	5.1	483	22	AAAB78975 C. glutamicum SRT
9	124.5	5.1	553	20	AAV38796 Neisseria meningit
10	123.5	5.1	604	21	AAV75522 Neisseria meningit
11	122.5	5.0	558	20	AAV38797 Neisseria meningit

12	121	5.0	386	21	AAW90863 Murine edg6 protei
13	119.5	4.9	604	21	AAV75523 Neisseria meningit
14	119	4.9	806	22	AAAG92136 C glutamicum prote
15	118.5	4.9	535	21	AAAB43136 Human ORFX ORF2900
16	118.5	4.9	535	21	AAV71061 Human membrane tra
17	118.5	4.9	535	22	AAAB72395 Human L-type amino
18	118	4.9	560	22	AAAG81108 Mycobacterium tube
19	117.5	4.8	3011	19	AAW77397 Hepatitis C virus
20	117.5	4.8	3011	19	AAW77398 Hepatitis C virus
21	117	4.8	513	20	AAV38786 Neisseria meningit
22	116	4.8	533	22	AAAB72393 Rat L-type amino a
23	115.5	4.8	505	18	AAW23275 Bordetella pertuss
24	115.5	4.8	604	21	AAV75521 Neisseria gonorrhe
25	115	4.7	3033	14	AAAR33214 NANBH virus strain
26	114.5	4.7	2261	10	AAAP90164 Peptide encoded by
27	114.5	4.7	2301	10	AAAP92047 Sequence encoded i
28	114.5	4.7	2435	13	AAAR25135 HCV polypeptide 1
29	114.5	4.7	2436	10	AAAP92050 Sequence encoded i
30	114.5	4.7	2436	10	AAAP90288 Peptide encoded by
31	114.5	4.7	2436	13	AAAR28582 HCV amino acid seq
32	114.5	4.7	2772	11	AAAR08123 Hepatitis C virus
33	114.5	4.7	2772	21	AAAB18540 Protein encoded by
34	114.5	4.7	2894	13	AAAR24440 Composite HCV HC-J
35	114.5	4.7	2894	16	AAAR70230 Composite hepatiti
36	114.5	4.7	2955	20	AAV14975 Amino acid sequenc
37	114.5	4.7	2955	21	AAAB18541 Polypeptide encode
38	114.5	4.7	3011	13	AAAR21519 Compiled HCV sequ
39	114.5	4.7	3011	14	AAAR31621 Hepatitis C virus
40	114.5	4.7	3011	17	AAAR90931 Hepatitis C virus
41	114.5	4.7	3011	18	AAAW34480 HCV polypeptid
42	114.5	4.7	3011	19	AAAW40038 HCV polypeptid
43	113.5	4.7	2984	22	AAAE00442 Hepatitis C virus
44	113.5	4.7	2984	22	AAAE00447 Hepatitis C virus
45	113.5	4.7	2984	22	AAAE00449 Hepatitis C virus

ALIGNMENTS

RESULT 1	
AAV22569	standard; Protein; 388 AA.
AC AAV22569;	
DT 17-NOV-1999	(first entry)
DE Bacterial general essential protein gep1713.	
XX	
DE Bacterial general essential protein gep1713.	
XX	
KW General essential protein; pathogenic bacteria; pathogen; inhibitor;	
KW bacterial growth.	
XX	
OS Streptococcus pneumoniae.	
XX	
PN WO9933871-A2.	
XX	
PD 08-JUL-1999.	
XX	
PF 30-DEC-1998;	98WO-US27918.
XX	
PR 31-DEC-1997;	97US-0070116.
XX	
PA (MILL-) MILLENNIUM PHARM INC.	
XX	
PI Youngman P, Filtz C, Murphy C, Guzman L;	
XX	
DR WPI; 1999-430230/36.	
XX	
DR N-PSDB; AA220360.	
XX	
PT Streptococcus pneumoniae general essential protein genes and proteins,	
PT useful for identification of antibacterial agents -	
XX	
PS Claim 1; Fig 13; 124pp; English.	

XX This sequence represents a *Streptococcus pneumoniae* general essential
 CC protein (GEP) protein of the invention. The genes encoding the GEP
 CC polypeptides are useful molecular tools for identifying similar genes in
 CC pathogenic microorganisms, such as pathogenic strains of *Bacillus*. In
 CC addition, the operons containing genes encoding GEP and the polypeptides
 CC themselves, are useful targets for identifying compounds that are
 CC inhibitors of the pathogens in which the GEP are expressed. Such
 CC inhibitors are useful for inhibiting bacterial growth by being
 CC bacteriostatic or bacteriocidal.
 XX
 SQ Sequence 388 AA;

Query Match	6.0%;	Score 146;	DB 20;	Length 388;
Best Local Similarity	21.2%;	Pred. No. 1.6e-06;		
Matches 98;	Conservative 76;	Mismatches 157;	Indels 132;	Gaps 23;

```

OY      22 FHRHFGSLRAWRASSOLLWMSSEALGFLLAVVYGSAFPVPSSALGLIAAIAAYWALLS   81
        | : : | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      6 fHeKtKy-----lsskell---llgjlIsI-----flp--fyllfvvvlclylIsI    46

OY      82 LTIDLR--QATPIHWLVLYWGVDALATGLSP--VRAAALVGLAKLTLYLVLFALAAR   136
        | : : | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      47 ftg-dmksilqkmgehpmlllfIsystvislaqnwmglvasvgmflftIfihy--qs     102

OY      137 VLRRNRLRSLLFSVV--VTTSLEFVSXGLEN----QWIYGEELATWDRNSVADETSR   188
        : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      103 lshukffrllqlfvlfgsvglSaafaslehfivkkfnayalfspmqvbnqrnaevttff--   160

OY      189 vSYLGPNP-----LIAAYLPTTAFAASAAGVWRGWLPKLLATAATGASSLCLI   238
        ||| : : | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db      161 -----npnyyglicfcfcimiafyllftctkln-----wlkvfcvla-----   195

OY      239 LTYSRGWLGFVAMIFWALLGLYNFQPRLPAPWRRWLFPPVLIGLVAVLLAV-----   292
        ||| | | | | | : | : | : | : | : | : | : | : | : | : | : |
Db      196 -----gfV-----nlfglnftq-----nrtafpaiagaillylfttknwkaF   233

OY      293 -----LGLEPLRVVLSIEVGREDDSSNNFRINWVLAVLOMIODRPWLGTGPNGTAFN   344
        : | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db      234 wlsigvfaiglsflfsdldgvrmgtlDssmeerisiwdagmalfkqpifwgegp-----   287

OY      345 LVYPLEYQQAARFTA--LSAYSVPLEVAAVEGULLGLTAFAMLLLVTAVTAVRQYSRLRRDR   401
        | : : | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      288 lty-mhsyprihapyehehahsiyiditIsygiVgtl---llvIssvaprtlmmdmsges   342

OY      402 NPOAFWMASLAGLAGMLGHGLEDTVLYRPEASTLMWLICIGAI 444
        : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      343 qkrpi-igilyisfltvavvhgidalf-----wiqsgfi 376

```

XX		
PE	27-JUL-1999;	99WO-GB02451.
XX		
PD	10-FEB-2000.	
XX		
AC	AAV81520;	
XX		
DI	24-MAY-2000	(first entry)
XX		
DE	Streptococcus pneumoniae type 4	protein sequence #20.
XX		
KW	Streptococcus pneumoniae; vaccine; screening; protein antigen;	
KW	antibacterial; antiinflammatory; meningitis; infection; diagnosis;	
KW	pneumococcal disease.	
XX		
OS	Streptococcus pneumoniae.	
XX		
PN	WO200006737-A2.	
XX		
XX		
XX		
ID	AAV81520 standard; Protein; 348 AA.	
AAV81520		
RESULT	2	

PR 27-JUL-1998; 98GB-0016337.
PR 19-MAR-1999; 99US-0125164.
XX
XX
PA (MICR-) MICROBIAL TECHNICS LTD.
XX
XX Gilbert CFG, Hansbro FM;
PI
XX
DR WPI; 2000-195300/17.

New Streptococcal protein, useful as a vaccine, for diagnosis of pneumococcal diseases and for screening agents capable of antagonizing or inhibiting expression of the protein -

Claim 1; Page 74; 108pp; English.

AA AAY81501 to AAY81679 represent specifically claimed protein sequences
CC isolated from *Streptococcus pneumoniae*. AAA05407 to AAA05590 represent
CC specifically claimed nucleotide sequences isolated from *S. pneumoniae*.
CC The sequences have antibacterial and antiinflammatory properties.
CC The protein sequences, and fragments of them, are useful as immunogens
CC and/or antigens. The nucleotide sequences can be used in vaccines and in
CC diagnostic assays. The proteins and nucleotides can be useful for the
CC detection and diagnosis of *S. pneumoniae*. The protein sequences are also
CC useful for screening an agent capable of antagonising, inhibiting or
CC interfering with the function or expression of the proteins in which the
CC agent is useful for treatment or prophylaxis of *S. pneumoniae* infection
CC and meningitis. AAA05591 to AAA05614 represent primers used in the
CC exemplification of the present invention.

Sequence 348 AA;

Query Match	5.9%;	Score 143.5;	DB 21;	Length 348;
Best Local Similarity	20.28;	Pred. No. 2.5e-06;		
Matches 78; Conservative	67;	Mismatches 135;	Indels 107;	Gaps 16;

QY	94	HWVLLVWGVDA	LATGUSP--VRAAALVGLAKITLVLLVFALAA	RYLRNPRLSL	ESVY	151				
		:: :	: :	:: :	:: :	:				
Db	11	hpmllfls	stvisllaqwmq	lvasvgmftl	fflhy--qslshk	ffrlllqfv1 67				
QY	152	---VITSLFVSV	YGLN---QMIYGV	EELATWDRNSVADFTSRVYSYLGNPN	-----	197				
		:: :	:	: :	: :	:				
Db	68	fgsvlsaa	faslshfqlvk	fnafslpmqvw	hqnraevtf	-----npnyyq11c 119				
QY	198	---LLAAYLVPT	TAFAA	AIIGVWRGWL	PKLLA	IAATGASSLCILLTYSRGWLG	FVAMI 253			
		:: :	:	:	:	:	:			
Db	120	cfcmla	fylfttck	ln-----w	lkvfcv	ia-----gfv	148			
QY	254	FVWALLGLW	FQPRLP	APWRWL	FPVVL	LG	LAVLLVAV-----IGLEPR 299			
		:: :	:	:	:: :	:	:			
Db	149	---nlfg	lnftq-----n	rtafpa	llagat	lyllftt	lknwka	fslsgvfa	iglsfl 198	
QY	300	VRYLSIFV	GREDDSSN	PRINW	LAVLQ	MIQDR	FWLG	IGPN--"IAFN	LVPLYQOAR	FTA 357
		:: :	:	:	:: :	:	:	:	:	:
Db	199	ssdlgv	rmgtld	ssmeer	isidag	malfkgn	pfwgegp	l	tymsypr	lhapyhe----- 253
QY	358	LSA	SVPLEVA	VEGGLGLTA	FAWMLLV	TA	VA	TRQVSR	LRDRNRQAF	WIMASLAG 417
		:: :	:	:	:: :	:	:	:	:	:
Db	254	-hahsly	idctilsy	glvgtl----	llvls	vavprl	mmdm	sgeq	krpi-igly	lsflcv 307
QY	418	MLGHGL	FDTVL	VRPEAST	LMWLC	IGAI 444				
		: :	:	:	:	:	:	:	:	:
Db	308	vavhg	ldla	lf-----w	iqsgfi 326					

DLT 3
9584
AAB79584 standard; Protein; 758 AA.

AC	AAB79584;
XX	
DT	30-APR-2001 (first entry)
XX	

DE Corynebacterium glutamicum SMP protein sequence SEQ ID NO:684.
 XX
 KW Corynebacterium glutamicum; carbon metabolism and energy production;
 KW SMP protein; sugar metabolism and oxidative phosphorylation protein;
 KW fine chemical production; organic acid; proteino-genic amino acid;
 KW nonproteino-genic amino acid; purine base; pyrimidine base; nucleoside;
 KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
 KW carbohydrate; aromatic compound; vitamin; cofactor; polyketide; enzyme;
 KW diagnosis; Corynebacterium diphtheriae; evolutionary study.
 XX
 OS Corynebacterium glutamicum.
 XX
 PN WO200100844-A2.
 XX
 PD 04-JAN-2001.
 XX
 PF 23-JUN-2000; 2000WO-IB00943.
 XX
 PR 25-JUN-1999; 99US-0141031.
 PR 08-JUL-1999; 99DE-1031412.
 PR 08-JUL-1999; 99DE-1031413.
 PR 08-JUL-1999; 99DE-1031419.
 PR 08-JUL-1999; 99DE-1031420.
 PR 08-JUL-1999; 99DE-1031424.
 PR 08-JUL-1999; 99DE-1031428.
 PR 08-JUL-1999; 99DE-1031431.
 PR 08-JUL-1999; 99DE-1031433.
 PR 08-JUL-1999; 99DE-1031434.
 PR 08-JUL-1999; 99DE-1031510.
 PR 08-JUL-1999; 99DE-1031562.
 PR 08-JUL-1999; 99DE-1031634.
 PR 09-JUL-1999; 99DE-1032180.
 PR 09-JUL-1999; 99DE-1032227.
 PR 09-JUL-1999; 99DE-1032230.
 PR 09-JUL-1999; 99US-0143208.
 PR 14-JUL-1999; 99DE-1032924.
 PR 14-JUL-1999; 99DE-1032973.
 PR 14-JUL-1999; 99DE-1033005.
 PR 27-AUG-1999; 99DE-1040765.
 PR 31-AUG-1999; 99US-0151572.
 PR 03-SEP-1999; 99DE-1042076.
 PR 03-SEP-1999; 99DE-1042079.
 PR 03-SEP-1999; 99DE-1042086.
 PR 03-SEP-1999; 99DE-1042087.
 PR 03-SEP-1999; 99DE-1042088.
 PR 03-SEP-1999; 99DE-1042095.
 PR 03-SEP-1999; 99DE-1042123.
 PR 03-SEP-1999; 99DE-1042125.
 XX
 PA (BADI) BASF AG.
 XX
 PI Pompejus M, Kroegeer B, Schroeder H, Zelder O, Haberhauer G;
 DR N-PSDB; AAF71701.
 DR
 XX
 PT New isolated Corynebacterium glutamicum nucleic acid encoding a sugar
 PT metabolism and oxidative phosphorylation protein for production or
 PT modulation of production of fine chemicals e.g. amino acids,
 PT carbohydrates or enzymes -
 XX
 PS Claim 20; Page 1109-1111; 1246pp; English.
 XX
 CC AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar
 CC metabolism and oxidative phosphorylation (SMP) proteins given in
 CC AAB79243 to AAB 79633 which are involved in carbon metabolism and
 CC energy production. The C. glutamicum SMP gene can be used in vectors
 CC (II) for expression in host cells and production or modulation of
 CC production of fine chemicals, such as, an organic acid, a proteino-genic
 CC or nonproteino-genic amino acid (preferred), a purine or pyrimidine base,
 CC a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty
 CC acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a
 CC cofactor, a polyketide, or an enzyme. The presence of (I) or SMP proteins

CC (III) encoded by them are used for diagnosing the presence or activity of
 CC Corynebacterium diphtheriae in a subject. (I), (II), (III) or host cells
 CC containing them are used to map genomes of organisms related to
 CC C. glutamicum, identify and localise C. glutamicum sequences of interest,
 CC in evolutionary studies, in determining SMP protein regions required
 CC for function, in modulating SMP protein activity, in modulating the
 CC metabolism of sugars, and in modulating high-energy molecule production
 CC in a cell (i.e. ATP, NADPH).
 XX
 SO Sequence 758 AA;
 Query Match 5.5%; Score 132.5; DB 22; Length 758;
 Best Local Similarity 22.5%; Pred. No. 0.0001;
 Matches 116; Conservative 67; Mismatches 188; Indels 145; Gaps 24;
 QY 20 SSFLHR-----LEGSLRAWRRASSQLLWSEALGGFLAVVYGSAPFPSSALGLGLAA 72
 Db |::||| ::||| :||| :||| |::|
 38 safhaaavvkaglyllrfslvfhdvawn-----wllliiv-----gmgtai 80
 QY 73 IAAWALLSTFDIDLKQATPIHMLVLLVWGVDALATGLSPVRAALV-----GLAKLTLY 127
 Db ::|||:| ::|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
 81 msayfa---vqkcdlkkltastvshlgvativgtqtpfalgaalvhtlshalfksslf 137
 QY 128 LLVFALARVLRNDRRLSLLFSVVVTTSLEFSVYVGLNQMITYGVEELATWDRNSVADFTS 187
 Db :|: : : :| | | :| | | :| | | :| | | :| | | :| | | :| | | :| | |
 138 mligvidhgt-grdirrlgflvkkmplfivsv-----lgaismas-----vp 180
 QY 188 RYVSYLGNPNLLAAYLVPTTAFSAATGVWRGWLPKLLAIAATGASSLCLILTYSRGWL 247
 Db : : : : :| | | :| | | :| | | :| | | :| | | :| | | :| | | :| | |
 181 pllgfvskegmi-----tafmdapig--nsyvlllvgaalga-----vltfysakl 226
 QY 248 GFVAMIFVWALIGLYWFQPR-----LPAPWRRLFPVVLGGLVAVLLVAVLGLEPLRV 302
 Db :|: :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
 227 -----vlgafvdgprdmshvkeavslwl-paalpgjmslplvlvlslfdapvsa 275
 QY 303 LSIFVGREDDSSNN-----FRINWLAVLQMTQDRPWLGI 336
 Db : : : : :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
 276 aatsaageaahmlalwhgintplllslglvlgvlfrkelw----klaetspf-pl 330
 QY 337 GPGNTAFN-LVYPLYQOARFTALSAVSYLE---VAVEGGLGLTAAFA----- 380
 Db | :| :| | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
 331 atgndilsmlyranllgkffgrmadmsprhlvslivllwalaafatihpsvqlapk 390
 QY 381 -----WLLVLTAVTAVRQVSRLRRDRNPQAFWIMASLAGMLGHGL-FDTVLY-RPEA 433
 Db :|: :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
 391 pyidrwldlplalaialsvfgllltnr-----rlsaavlvgtvgvgsfqllllgapdv 444
 QY 434 STLWMLCTGAIASF-----WQPOPSKOLPPEAHS 463
 Db : :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
 445 altqflvegllvvvimvvrhqpafnkfrkpsrrrs 480
 RESULT 4
 AAG90046
 ID AAG90046 standard; Protein; 962 AA.
 XX
 AC AAG90046;
 XX
 DT 26-SEP-2001 (first entry)
 XX
 DE C glutamicum protein fragment SEQ ID NO: 3800.
 XX
 KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
 KW organic acid synthesis.
 XX
 OS Corynebacterium glutamicum.
 XX
 PN EPI108790-A2.
 PD 20-JUN-2001.
 PF 18-DEC-2000; 2000EP-0127688.

Db 12 aligivlsvltfwlfaqstlnigpdmadlqmsdgtlmnlavvaalfcqtf----- 62

QY 171 EELATWDRNSVADFTSRV-YSYLGN-PNLLAAYLVPTTAFSAAGVWRGWLKLLAIA 228

Db 63 -----ivaaggiadvfgvrirmmignlnilgslilatattslatqmvltgrvlglaaa 117

QY 229 ATGASSLCILITV-----SRGGW-LGFVAMI-----FWW-----AL 258

Db 118 aimsasialvktlywlgtdrqravsiwsgwsgtgfcalfaglvvaaspfgwrgifalcal 177

QY 259 LGLYWFQPRLLPAPWRRLFPVVLGGL-----VAVLLVAVLGLE----- 296

Db 178 vsivaialtrhlpestr---paqsigmhlwsglivalavlslelffitqgeslgwthmt 234

QY 297 ----PLRVRLSIFVGREDDSSNNFRINWV-----LA 323

Db 235 wtlavsltlflavfvie-----riaswpldlnl fkdhafsgatitnflmsatgqvva 288

QY 324 VLQMIQDRPWLIGIP-----GNTAFNLVYPLYQOARFTALSAYSVPLEVAVEGGLLG 375

Db 289 vvmwvqgmw-gvaptisgltsigfaafvlflirvgekamqkvgara-----viltagilv 343

QY 376 LTAFAWLLLVTAVTAVROVSRLRDRNPQATWMAASLAG--LAGMIGHGIEDTVLYRPEA 433

Db 344 atata-limits-----estylivislagsfslg-lylgilfatpvtal 387

OY 434 STL 436

Db 388 gtl 390

RESULT 9

ID AAY38796 standard; Protein; 553 AA.

XX AAY38796;

DT - 08-OCT-1999 (first entry)

XX

DE Neisseria meningitidis antigen encoded by ORF141.

XX

KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine; treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.

XX

OS Neisseria meningitidis.

XX

PN WO9924578-A2.

XX

PD 20-MAY-1999.

XX

PF 09-OCT-1998; 98WO-IB01665.

XX

PR 01-SEP-1998; 98GB-0019016.

PR 06-NOV-1997; 97GB-0023516.

PR 14-NOV-1997; 97GB-0024190.

PR 18-NOV-1997; 97GB-0024386.

PR 27-NOV-1997; 97GB-0025158.

PR 10-DEC-1997; 97GB-0026147.

PR 14-JAN-1998; 98GB-0000759.

XX

PA (CHIR-) CHIRON SPA.

XX

PI Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;

XX

DR MPI; 1999-327407/27.

DR N-PSDB; AA212229.

XX

PT Proteins from Neisseria meningitidis and N. gonorrhoeae useful for diagnosis, treatment and prevention of infection

XX

PS Claim 4; Page 338-339; 524pp; English.

XX

CC Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis

CC and N. gonorrhoeae antigenic proteins. They are encoded by open

CC reading frames (ORFs) AA21972-212358. The antigenic proteins,

CC their fragments, their nucleic acids and antibodies are used for

CC diagnosis, prevention (as vaccines) or treatment of Neisseria

CC infections, such as meningitis, septicaemia and gonorrhea. Both

CC organisms are closely related. Fragments of the nucleic acids

CC are useful as hybridisation probes and antisense reagents.

XX

Sequence 553 AA;

Query Match 5.1%; Score 124.5; DB 20; Length 553;

Best Local Similarity 24.8%; Pred. No. 0.00043;

Matches 109; Conservative 43; Mismatches 141; Indels 147; Gaps 23;

QY 79 LSLTFDIDLROATPIH---WLVLL-----YMGVDALATGLSP 112

Db 1 mltlytpdpdarpakthekpwlilllmatawlpvgvfshdlwmpdepaytavealagsptp 60

QY 113 VRAAALVGLAKL---TLVLLVFALAAARVLRNP-----RLRSLFSVVITSLFVSV 160

Db 61 l-vahlfgqtdfgiprvlw-aaafkhlispwaadsydaarfagvffavlgltscgfaag 118

QY 161 YGLNWIYGVEEELATWDRNSVADFTSRVSYLGNPNLL--AAYLVPTTAFSAAGVWR 218

Db 119 fnflgrbhg-----rsvvlllgciglvpahftlnpaafaaaglvln 162

QY 219 GW-LPKLLATAA-----TGASSLCILITVSRGGWLGFAVAMIFWALLGLYWFQPRLPAPW 272

Db 163 gyslarrrviaasfligtgtwlmslaaayp-----aafalmplpvlmffr-----pw 210

QY 273 --RRWLFPPVVLGGLVAVLLVAVLGLEPLRVRLSIFVGREDDSSNNFRINWMLAVLQMIQD 330

Db 211 qsrllmltavaslafalplmtvyp-----llaktqpal 245

QY 331 RPWL-----GIGPGNTAFNLVYPLYQOARFTALSAYSVPLEV-----AVEGG 372

Db 246 aqwldghvftgtfgvrhvqtatfsflfyllknlwf-alpa--lplawtwcrtlrfsdwtg 302

QY 373 LLGLTAFAMLLLVTAVTAVROVSRLRDRNPQAF-----WMAASLAGLAGMIGHGIEDTV 427

Db 303 ilgv---vwmlavllvllav-----npqrfgdnlvwllippla---lfgaaqls- 344

QY 428 LYRPEASTLWMLCIGAIASF 447

Db 345 lrrgaafvwmfglmafglf 364

RESULT 10

AAY75522

ID AAY75522 standard; Protein; 604 AA.

XX AAY75522;

DT 21-MAR-2000 (first entry)

XX

DE Neisseria meningitidis ORF 738 protein sequence SEQ ID NO:2518.

XX

KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine; antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia; antibacterial; gene therapy.

XX

OS Neisseria meningitidis.

XX

PN WO9957280-A2.

XX

PD 11-NOV-1999.

XX

PF 30-APR-1999; 99WO-US09346.

PR 01-MAY-1998; 98US-0083758.

PR 31-JUL-1998; 98US-0094869.

PR 02-SEP-1998; 98US-0098994.

Db 176 fligt-----gwtlmslaaaypaafalmjlp-----vlmff----- 207
QY 262 YWFQPLPAPW--RRWLFPVVLGGLVAVLLVAVLGLLEPLRVRLSIFVGRDSSNNFRIN 319
Db 208 -----rwpqsrrlmltavaslafalplmtvpl----- 235
QY 320 VWLAVLQMIODRPWL-----GIGPGNTAFNLVYPLYQQAFTALSAYSPLLEV--- 367
Db 236 -llaktqpalfaqwlddhvfgtfgvrhigtatfslfyllknlwf-aipla--lplavwtv 291
QY 368 -----AVEGGLGLTAFAMWLLVTAVTAVKQVSRLRDRNPQAF-----WLMASLAGLA 416
Db 292 crrlfsfdwglgv-----vwmilavlllav-----npqrfqdnlvwllppla--- 335
QY 417 GMLGHGLFDTVLYRPEASTLWMLCIGATASE 447
Db 336 -lfgaaqls-lrrgaaafvnmfglmafglf 364

RESULT 12

AAW90863
ID AAW90863 standard; Protein: 386 AA.

XX AAW90863;

DT 29-JUN-2000 (first entry)

DE Murine edg6 protein.

KW edg6; murine; G-coupled receptor; endothelial differentiation gene;

KW antinflammatory; immunomodulatory; antimicrobial; antiallergic;

KW cytosolic; gene therapy; inflammation; autoimmune disease; allergy;

KW tumor; leukemia; lymphoma.

OS Mus sp.

PN DE19846979-A1.

XX 23-MAR-2000.

PF 13-OCT-1998; 98DE-1046979.

XX 11-SEP-1998; 98DE-1043240.

PA (DELB-) DELBRUECK CENT MOLEKULARE MEDIZIN MAX.

PI Graeler M, Bernhardt G, Lipp M;

DR WPI; 2000-258069/23.

DR N-PSDB; AAX82790.

PS Claim 2; Page 6; 12pp; German.

CC This invention describes novel human and murine G-coupled receptors EDG
CC (endothelial differentiation gene) 6 (I and II). The products of the
CC invention have antinflammatory, immunomodulatory, antimicrobial,
CC antiallergic and cytosolic activity. (I) and (II) are involved in signal
CC transduction. (I), (II) and their fragments, variants and mutants or
CC binding partners, are used therapeutically to modulate the function of
CC blood and body cells, particularly for inhibition of acute and chronic
CC inflammation and to raise specific antibodies against them. They are used
CC as a source of diagnostic oligonucleotides and for gene therapy.
CC Antibodies against (I) and (II) are useful for diagnosis and optionally
CC when coupled to therapeutic agents, toxins or other antibodies, to
CC modulate immune and inflammatory responses for example immunological
CC defects such as inflammation, infection, autoimmune diseases, allergy,
CC tumors, leukemia and lymphoma. This sequence represents the murine EDG6
CC protein described in the method of the invention.

SO Sequence 386 AA;

Query Match 5.0%; Score 121; DB 21; Length 386;
Best Local Similarity 27.3%; Pared. No. 0.0006;
Matches 117; Conservative 33; Mismatches 98; Indels 180; Gaps 28;

QY 1 MTWQTLFAHYPOQWGHSSFLHRLFGSLRAWASSQLLV-----WSEALG 47
Db 3 lstwstlv-----tpes-----chrlaas-----ghslilivhyhnsgrlasrgsedgg 47
QY 48 GFLLAVVYGSAPFVPSALG-----LGLAIA-----ATWALLSTIDIDLQAT 91
Db 48 g--lgmrlg-----psvaagclvlenamvlaaiaiymsrrwyyclnltlsd----- 96
QY 92 PIHWLVLYWGDALATG-----LSPYRAAALVGLAKLTLVLYFAALAAARVLRNPRRLSL 146
Db 97 ---ltglayvvnvlsgtrtfglspvhwflreg-----llfmaiaastf-----s1 139
QY 147 LFSVWVITSLFVSVYGLNQWITYGVEELATWVD-RNSVADFTSGRYSYLGPNLLAAYLVP 205
Db 140 lft-----agerfatmvrvaesgatktsrygciglcwllaa1-- 178
QY 206 TTAFSAALIGWNRGWLPKLLAIAATGASSLC-----LILTVSRGGLGVAMIF---V 255
Db 179 -----gllpll-----gmcvcafpccssllplysk-gyvlfcvvfa11 218
QY 256 WALLGLYWF-----QPRLPA--PWRRLW-----FPVVLGGLVAVLLVAVL 293
Db 219 valslsygaifrvrangqksprparrrksrllntvmlvafvvcwgpilfgllladif 278
QY 294 G-----LEPLR-----VRVLSIFVGRDSSNNFRIN-----VWLAVLQMI-QDRPW 333
Db 279 gsnvwageylrgmdwila1avf-----nsainplysfrsrevgravlaflcgc1w 330
QY 334 LGI-GPGN 340
Db 331 lglrpgd 338

RESULT 13

AA75523
ID AAY75523 standard; Protein: 604 AA.

XX AAY75523;

DT 21-MAR-2000 (first entry)

DE Neisseria meningitidis ORF 738 protein sequence SEQ ID NO:2520.

KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;

KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;

KW antibacterial; gene therapy.

OS Neisseria meningitidis.

PN WO9957280-A2.

XX 11-NOV-1999.

PF 30-APR-1999; 99WO-US09346.

XX 01-MAY-1998; 98US-0083758.

PR 31-JUL-1998; 98US-0094869.

PR 02-SEP-1998; 98US-0098994.

PR 02-SEP-1998; 98US-0099062.

PR 09-OCT-1998; 98US-0103749.

PR 09-OCT-1998; 98US-0103794.

PR 25-FEB-1999; 99US-0121528.

PA (CHIR) CHIRON CORP.

PA (GENO-) INST GENOMIC RES.

Db 254 rignlsyangsvrgtlstrmhqegdlvcklwlavavlclaaavamwr----- 302
QY 230 TGASSLCLILTYSRGGWIGFVAMIFWALLGLYWFQRLPAWR--WLFPPVIG---- 282
Db 303 -----vvragnpygavmlnsllal-----cspvswshwvwlplalglas 345
QY 283 -----GLVAVL-LVAVLGLLEPLRVRLSIFVGRDS-----SNFRINVWL 322
Db 346 awngrrtapglaataqylalttptmfitftwmpydsesypfwplllqpsgnavvvvi 405
QY 323 AVLQMIQDRPWLIGIPGNTAFN-----LVYPLYQQARF-----TALS 359
Db 406 allivainptv-lgsunkavsgqaeksspallvllaiaifylfaniwfkgnnqkali 464
QY 360 AYSVPLEVAVEGGLGLTAFWLLVTAVTAVRQVSRLRDRMPQAFWLMASLAGL 419
Db 465 qy--plq-lmeg--rgltldgellifeaasnqlvs-----lwiigalnaia--- 506
QY 420 GHGLFDVLYRPEASTLWML 439
Db 507 -----latlwlfl 514

RESULT 15

AAB43136
ID AAB43136 standard; Protein; 535 AA.

AC AAB43136;

DT 08-FEB-2001 (first entry)

DE Human ORFX ORF2900 polypeptide sequence SEQ ID NO:5800.

Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
hypotensive; dermatological; immunosuppressive; antiinflammatory;
antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
neurodegenerative disorder; osteoarthritis; graft vs host disease;
cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
cholesterol ester storage; systemic lupus erythematosus; infection;
severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
bone damage; cartilage damage; antiinflammatory disease; coagulation;
thrombosis; contraceptive.

OS Homo sapiens.

PN WO200058473-A2.

PD 05-OCT-2000.

PF 31-MAR-2000; 2000WO-US08621.

PR 31-MAR-1999; 99US-0127607.

PR 02-APR-1999; 99US-0127636.

PR 05-APR-1999; 99US-0127728.

PR 30-MAR-2000; 2000US-0540763.

PA (CURA-) CURAGEN CORP.

PI Shimkets RA, Leach M;

DR WPI; 2000-602362/57.

DR N-PSDB; AAC77345.

PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
PS Claim 11; Page 4966-4968; 5507pp; English.

XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antithyroid; and antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 535 AA;

Query Match 4.9%; Score 118.5; DB 21; Length 535;
Best Local Similarity 22.6%; Pred. No. 0.0017;
Matches 89; Conservative 56; Mismatches 123; Indels 125; Gaps 19;

QY 1 MTWQTLPFAHYQ---PQQGHSSFLHRLFGS---LRAMRASSQLLWSEAL-----CG 48
Db 85 itvgalcyaelgvtipksggdyvkdifgylagflrlwlaavlvyptnqavialtfsn 144
QY 49 FLAAVVGSAPEVPSSALGLIAIAAYWALLSLTDID---LRQATPIHML----- 96
Db 145 yvlqplfpfc-fpesqlrl-laef-----clllltwncssvrvatrvqdlftagklial 198
QY 97 -VLLYWGVDALATG---LSPVRA-----AALVGLA----- 122
Db 199 allimglvqickgeyfwlepknafentqepdiglvalafiqsfaygwnflnyvteel 258
QY 123 -----KLTLYLVFALAAARV-LRNPRLRSLFSVVVITSLFVSVYGLNQW 166
Db 259 vdpyknlpraifisiplvtfvyanvayvtamp--qellasnnavavltfgekllygmaw 316
QY 167 IYGV-ELATWVDNRNSVADFTSRVSYLGNPNLLAAVLYPTTAFSAAAGVWRGWLPKL 225
Db 317 impisvalstfgvngslftsrllf-----fagareghlpsvl 354
QY 226 -----AIATGASSLCLILT--YSRGWLGFWAMIF---VWALLGLYWFOP 266
Db 355 amihvkrctripallftcistllmlvtsdmtyllinyvgfynylyfgyvtvagqivlrwkkp 414
QY 267 RLPAWR-RWLEPPVVLGLVAVLLVAVLGLLEPL 298
Db 415 diprplknllfpilyllfwalfllvslwsepv 447

Search completed: January 10, 2002, 23:04:03
Job time: 2290 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2002, 23:04:53 ; Search time 33.48 Seconds
(without alignments)
1062.530 Million cell updates/sec

Title: US-09-887-038-3
Perfect score: 2426
Sequence: 1 MTVMQTLTFAHYQPOQWGH...WQPPSKQLPPFAHSDEKM 467

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1132	46.7	474	2	S75464 hypothetical prote
2	206	8.5	438	2	A82262 conserved hypothet
3	201.5	8.3	387	2	S74522 hypothetical prote
4	195	8.0	443	2	D82847 membrane protein x
5	180	7.4	478	2	A83368 hypothetical prote
6	169.5	7.0	384	2	H75597 hypothetical prote
7	166.5	6.9	597	2	H82328 hypothetical prote
8	159	6.6	861	2	S77086 hypothetical prote
9	151.5	6.2	401	2	D83022 hypothetical prote
10	151.5	6.2	461	2	G83602 hypothetical prote
11	139.5	5.8	388	2	G75375 hypothetical prote
12	138.5	5.7	402	2	E86038 probable LPS biosy
13	138	5.7	515	2	G75267 ABC transporter, p
14	137.5	5.7	443	2	F83759 cytochrome d (bd-t
15	137.5	5.7	474	2	H75389 NADH dehydrogenase
16	134.5	5.5	660	2	A64739 ferrichrome-iron t
17	134	5.5	335	2	G82224 probable vitamin B
18	134	5.5	402	2	A83398 probable MPS trans
19	133.5	5.5	585	2	S74477 hypothetical prote
20	132.5	5.5	660	2	E85499 hypothetical prote
21	131	5.4	524	2	T02499 hypothetical prote
22	130	5.4	649	2	F83779 iron (III) dicitra
23	129	5.3	487	2	B72755 probable antibioti
24	128.5	5.3	530	2	G83631 cytochrome-c oxida
25	128.5	5.3	566	2	H84203 phosphate ABC tran
26	127.5	5.3	437	2	D83329 conserved hypothet
27	127	5.2	591	2	F70871 probable membrane
28	127	5.2	709	2	F75584 hypothetical prote
29	126	5.2	524	2	T03507 cephamycin export

30	126	5.2	776	2	A84178 cationic amino aci
31	125.5	5.2	404	2	G83322 hypothetical prote
32	125.5	5.2	787	2	H70374 NADH dehydrogenase
33	125	5.2	403	2	C83422 nitrate transporte
34	125	5.2	486	2	F83315 NADH dehydrogenase
35	124.5	5.1	328	2	D82411 conserved hypothet
36	124.5	5.1	553	2	C81024 hypothetical prote
37	124.5	5.1	643	2	G75599 probable drug tran
38	124	5.1	363	2	A72771 hypothetical prote
39	123.5	5.1	382	2	D75541 sodium/proton anti
40	123.5	5.1	500	2	F83418 NADH-ubiquinone ox
41	123.5	5.1	501	2	G82822 probable permease
42	123.5	5.1	539	2	H82994 hypothetical prote
43	123.5	5.1	404	2	A81181 O-antigen ligase c
44	122.5	5.0	404	2	B41317 probable inner mem
45	122.5	5.0	558	2	F81967

ALIGNMENTS

RESULT	1	Query Match	46.7%	Score 1132	DB 2	Length 474
S75464	hypothetical protein slr1515 - Synechocystis sp. (strain PCC 6803)	Best Local Similarity 47.5%; Pred. No. 2.8e-73;				
C:Species: Synechocystis sp.		Matches 223; Conservative 79; Mismatches 163; Indels 4; Gaps 2;				
A:Variety: PCC 6803						
C:Date: 25-Apr-1997	#sequence_revision 25-Apr-1997 #text_change 08-Oct-1999					
C:Accession: S75464						
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Miyajima, O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yas						
DNA Res. 3, 109-136, 1996						
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys						
S.						
A:Reference number: S74322; MUID:97061201						
A:Accession: S75464						
A>Status: preliminary						
A:Molecule type: DNA						
A:Residues: 1-474 <KAN>						
A:Cross-references: EMBL:D90911; GB:AB001339; NID:g1653083; PIDN:BAA18025.1; PID:d101						
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996						
QY	1 MTVMQTLTFAHYQPOQWGHSSPLRLFGSLRAWRASSQLVWSEALGFLAVYGSAPF 60					
DB	5 ISINSLMFGGFSPPQEWGRGVLRHVLWGQSWIQASVLMRPHFALGTALVAIIFIAAPF 64					
QY	61 VPSSALGLGLAIAAYWALSLTIDLRQATPIHMLVLLVWGVDAIATGSPVRAALVG 120					
DB	65 TSTMLGIFMLLCGAFWALTFADQPKGLTFPIHVLVAVWCISAIAYGFSVKMAASG 124					
QY	121 IAKLTVLLVLFALAAARVLRNBRSLFSVVVITSLFVSYYGLNOMIYGVLELATWVDRN 180					
DB	125 IAKLTANLCLFLAARLLQNRQWLNRLVTVLLVGLVGSYGLRQQVDGVEGLATWNDPT 184					
QY	181 SVADETSRVYSYIGNPNLLAAYLVPTTAFSAAGICVWRGWLPKLIAATGASSLCILIT 240					
DB	185 SFLAQATRVYSFLGNPNLLAAYLVPMTGLSLALVYWRWPKLLGATMVIVNLCLFTT 244					
QY	241 YSRGCMGLFVAMIFWALGLTWFOPLPAPWRKMLFPVVLGGLVAVLL--VAVLGLEPL 298					
DB	245 QSRGWLAVLALGATFLALCYFWWLPQLPKFWQKSLPLAIA--VAVILGGALIAVEPI 302					
QY	299 RYRVLSIFVGRDSSNNFRINWVLAFLQMDRPMGLIGPGNTAFNLVYPLVQARFTAL 358					
DB	303 RLRAWSIFAGREDSSNNFRINWEGVKAMIRARPIIGPGNEAFNQIYPYMRPFTAL 362					
QY	359 SAISVPLEVAVEGGLIGTAFAMLLVTAVTAARQVSRRLRDRNPQAFWMAASLAGLAGM 418					
DB	363 SAISYILEILVETGVGFTCMMLWLVAVTLGKGVLELVKRCROTLPAREGIWINGALAAIIGL 422					

submitted to GenBank, June 2000

A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigrychado, M.A.; Madelira, A.M.B.N.; Madelira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.; Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira M.; Tshnako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF0103

Query Match	8.0%;	Score 195;	DB 2;	Length 443;
Best Local Similarity	25.2%;	Pred. No. 1e-06;		
Matches 107;	Conservative 47;	Mismatches 158;	Indels 112;	Gaps 21;

OY	56	GSAPFVPSSALIGLGLAAIAAYW-----ALLSLTDIDIQATPIHMLVLLYWGDALA	107
Dd	63	GGTQLLDSPAWALTSTVLFMAYWLQVVSLFGALELESLSRKYATDLRYLPFWM--LCAIA	120
OY	108	TGLSPVRAAALVGKLKTLTYLLVFALAAARVLNRNRLRSLLFSVVITSLFVSYYGLNOMI	167
Dd	121	VGSPERRERTEFKGLAVIGLTWTIDLALQAAMLHT-----SPLFWSLNQLRQAI	167
OY	168	YGVEELATWVDNRNSVADFTSRVYSYLGNPNLLAAYLVPT-----AFSAALGVWRGWL	221
Dd	168	SGYGFCS-----ROMI--LADRLSGTGPONLKIGQTLATLSPFLVLFAMQRGLALW----	218
OY	222	PKLTAIAATGASSLCILITYSRGWLGF--VAMIFVMALIGLYWFQPRLPAPWRBLLFPV	279
Dd	219	--LIVVYVLG---IVLLAGSRASMITTAYALIMLSGRVLCI-----RLL---	259
OY	280	VLGLIYAVLLVAVLGLEPL---RV-RVLSIFVGR---EDSSNNFRINWMLAYLOMIODRP	332
Dd	260	GMGALFLPLALAVIAFSPTRERIDRTAAVEFADHGAGVDQALSGRSQIWQAALCMIOAHNP	319
OY*	333	WLGIGPNTAFNLVYP-----LYOQARFTALSAYSVPLEVAVEGGLGL-----TA	378
Dd	320	LSGVGV--RGFRDAYPACNPPIPERIPAWGAGPALHAHOIVLEILSETGITGLELMIIAGAA	377
OY	379	FAMLLLTAVTAVRQVSRLRDRNDPOA-----FW--LNASLAGL-	415
Dd	378	MAW-----RAWRYATAARDRARAPAMISLTAIVEPENHTLAIFYSSFWMALMLLAGLY	430
OY	416	AGML 419	
Dd	431	AGAL 434	

RESULT 5
A83368
hypothetical protein PA2240 [imported] - Pseudomonas aeruginosa (strain PA01)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: A83368
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warener, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A;Reference number: A82950; MUID:20437337
A;Accession: A83368
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-478 <STO>
A;Cross-references: GB:AE004649; GB:AE004091; NID:g9948246; PIDN:AAG05628.1; GSPDB:GN001
C;Experimental source: strain PA01
C;Genetics:
;Gene: PA2240

Query Match	7.4%;	Score 180;	DB 2;	length 478;
Best Local Similarity	23.0%;	Pred. No. 1.3e-05;		
Matches 124; Conservative	67;	Mismatches 159;	Indels 188;	Gaps 29;

```

QY 48 GFLLAIVYGSA-----PFVPPSSALGIGIAIAA-----YWALL-----SL 82
      | : | | : | | | : | : | : | | | : | | |
Db 4 GNLVAVVFGLAFGALIMLSPAKAGAAM-VGLAAAVTILRFPFWGLLEFALVATFMPYST 62
      | : | : | : | : | : | : | : | : | : | : | : |
QY 83 TPDIDRQATPIHMLVLLYGV--DALATGISPVRAALVGLAKTLYLLVFA----- 132
      : : | : | : : | | : | | | : : | : | : |
Db 63 VNWGIR-STVSEAILALTWGAVALMHIFLSRLPPAPALRYRSTERMILMLFTVLPFYVG 121
      : : | : | : : | | : | | | : : | : | : |
QY 133 -----LAARVL-----RNPRLSLFSVVITSLFVSZY- 161
      | | | : | : | : | : | : | : | : | : |
Db 122 QVSIKAEASGLSNMLRMLNLSIVFIAGRLVERKNRETLLVIALLEGLTAMLLMSIAVFI 181
      | : | : | : | : | : | : | : | : | : | : |
QY 162 -----GLNQWITYGVEELATWDRNSVADFTSRVSYLGNPNL--LA 200
      | : | : | : | : | : | : | : | : | : |
Db 182 RYRSASGMAPILAMFNYSANLDTLKFGLLEAL-----SSRRGSPWMHPNATGIGIM 229
      | : | : | : | : | : | : | : | : | : |
QY 201 AYLVPPTTAFSAAGVWGRWMLPKLLAIAATGASSLCLILTYSRGCVLGFVAMIFWALIG 260
      | : | : | | | | | : | : | : | : | : | : | : |
Db 230 ALLPLPACFYGVANGWRGLG--LAVALVIGAAA--ILLASSRGMLSLAVVL-F-WMSLR 284
      | : | : | : | : | : | : | : | : | : | : |
QY 261 LWFQPRLLPAPWRRLFPVVLG-GLVAVLLVAVLGLLEPLRVRVLSIFVGRDSSNNFRIN 319
      : | : : | : | : | : | : | : | : | : | : |
Db 285 KVPYTGRL-----LLGVGLVVVL--VLSYPLQERLATIF-SPQNASTEYVRFD 330
      : | : : | : | : | : | : | : | : | : | : |
QY 320 VMLAVLQMIQDRPNLIG-----PG-----NTAFNLVYPLYQQARFTALASYSVPL 365
      : : | : | : | | | | | : | : | : | : | : |
Db 331 EYRMFPKAVARYP-LGIGFKVDPVPDGLLGISNMLNPMY----- 371
      : : | : | : | : | : | : | : | : | : | : |
QY 366 EVAVEGGLLGLTAFAMLLLVTAVTAVRQVSRLRRDNP-----QAFWLMAISLAGLGM 419
      : : | : | : | | | : | : | : | : | : | : |
Db 372 KVGL-GGMLLFIAVTW-----RWMREARPEKGPIRLTRDNATWL-GSTVGILAL 419
      : : | : | : | : | : | : | : | : | : | : |
QY 420 GHGLFDTVL-YRPEASTLWMLCIG-----AIASFQWQOP-----SKQLPPEAE 461
      | | | : | : | : | : | : | : | : | : | : |
Db 420 VSGLFDDHYESFAVVMIGLEFWLLVGINLLEARRLPPERQOPPAVGYYKLKRLERGAE 477
      | | | : | : | : | : | : | : | : | : | : |

```

```

RESULT      6
H75597
hypothetical protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: H75597
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
  M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
  S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
  Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896
A:Accession: H75597
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-384 <WHI>
A:Cross-references: GB:AE001862; GB:AE001825; NID:g6460468; PIDN:AAF12372.1; PID:g646
A:Experimental source: strain R1
C:Genetics:
A:Gene: DRA0046
A:Map position: 2

```

	Query Match	7.08;	Score 169.5;	DB 2;	Length 384;	
	Best Local Similarity	26.7%;	Pred. No. 5.7e-05;			
	Matches 113; Conservative	56;	Mismatches 158;	Indels 97;	Gaps 24;	
<hr/>						
Oy	50 LLAVVYGSAPFVPSASGIGLAAITAAYWALLSLTDIDLRQATPIHMLVLLYGVDALATG	109				
	: : :	:	:	:	:	
Dd	21 LIALV---PVFP-----LYLALACLGSLRTP-----CAAR---GVLEFFFAATOLIAA	64				

```

QY      110 LSPVRAALVGLAKL-TLYLLVFALAAVLNR- LRSLLFSVVITSLFVSYYGLNOMI 167
      |      |||      ||      |      ||      |      ||      |      ||      |
Db      65 LFTPAPLLSVGLAAARTLLIIMAVAGVYLRDSRHLRPLLMGOLI--FVTA-----WI 116

QY      168 YGVEELATWDRNSVADFTSRV--YSTLGNPULLAAYLVPTTAFSAAGVWRGMLPKL 224
      |      |      |      |      |      |      |      |      |      |      |
Db      117 Y--TLST---QGAVGVQERLGHPIYYIVSLGLVAVVALMIVMFWRGAAMWR-WPAGL 168

QY      225 LAIAATGASSICLILYSRGWLGFVAMIFWVALGLYWFQPRLLPAPWRW-LPPVYLG 283
      ||      |      |      |      |      |      |      |      |      |      |
Db      169 LALA-----TFAAAGSRGPLLALGVGSLAALAFGGQR---RWVMDPAVLVM 213

QY      284 LVAVLLVAV-LGLEPLRYRVLSIFVGREDSNNFRINWLAVLQMIQDRPWLIGPGNTA 342
      |      |      |      |      |      |      |      |      |      |      |
Db      214 LAAFATTSLNVPFKPLDRLLNDQTSGRE-----YVQDAVSGWETSPLGVGPPYQGG 265

QY      343 FNLVY-----PLXQQ-----ARFTA--LSAVSVPLEVAVEGGLGTAFAWL 382
      |      |      |      |      |      |      |      |      |      |      |
Db      266 PLYTYLFFKDCQCLTFYLQNKIECPQQLSRWSSVWLIAHNAMLHMLLESIGILGSG---- 321

QY      383 LLVTAVTAVRQVSRLRRDRNPQAFWLMASLAGMLGHGLEDVLY--RPEASTLWMLC 440
      ||      |      |      |      |      |      |      |      |      |      |
Db      322 LLAIMVYALWRAIQI---GDP-----FTLAVLYGFTAMNVVDVYIAVPSPHFSELMWVC 372

QY      441 IGAII 444
      |      |
Db      373 VGLV 376

```

```

RESULT      7
H82328
hypotheoretical protein VC0393 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
C;Species: Vibrio cholerae
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C;Accession: H82328
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, M.D.; Yamathavan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, F.
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A;Reference number: AB2035; MUID:20406833
A;Accession: H82328
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-597 <HEI>
A;Cross-references: GB:AE004127; GB:AE003852; NID:g9654808; PIDN:AAF93566.1; GSPDB:GN001
A;Experimental source: serogroup O1; strain N16961; biotype El Tor
C;Genetics:
A;Gene: VC0393
A;Map position: 1

```

	Query Match	6.9%	Score 166.5;	DB 2;	length 597;
	Best Local Similarity	24.3%;	Pred. No. 0.00015;		
	Matches 109;	Conservative 60;	Mismatches 185;	Indels 95;	Gaps 22;
QY	49 FLAAVYGSAP--FVPS-SALGLGIAIAAYWALLSLT-----DIDLRQATPIHMLVL 99 : : : : : :				
Dd	29 FALAAYYLLAMHFMPNPGAGLALSFNNTVMATSTLATIGLYQLANNQALRYSKLTIG 88				
QY	100 YMGVDALATGLSPV-----RAALVGL-AKLTYLLVFALARVLRNPRLSLFFS 149 : : : : : :				
Dd	89 LWLSCALLT-LPITYSNADWFHTLCRLVGIMAGFTLFVLVLOQF-RFSNKHKQRLLMF- 143				
QY	150 VVVITSLFVSVYGLNQWITYGVEELATWVDNRNSVADF--TSRVSYSLGNPNLLAAYLVPT 206 ::: : : : : : : : :				
Dd	144 -ILLGSIVLAIVIGLIQYF-----WLEPGNPFGYDTKANRPYGIFQQPNVMASFLATG 194				
QY	207 TAFSAALIGVWRGWL-----PKLLAIATG-ASSICLICILTYSRGGWLGFWAMI 253 : : : : : : :				
Dd	195 LVLS-----GYLLARQPEKYDSRMSKIGILYATPLLTAPLLVVIASRTGMLASLISI 246				
QY	254 FVVALIGLYWFQRPRLPAPWRRWLFPVVLGGLVAVLLVAVLGLGLEPLRVRVLSIFVGREDSS 313				

```

Db      247 ---ALVAPYLYPEASRRRFTLMLAITLG-ICAGFTAMYG-----QGTSGFVANKADL 295
QY      314 NNERINWLAVLQMIQDRPWLIGIGPGNTAFNLVPLYQQAARTALSAYSVPLEV----- 367
Db      296 ESPRRTFEPQITDMLIEKPFGTGYGK--FEAQYLLYTARQHQINSYYPAGLASMDHPHN 353
QY      368 -----AVEGILGUTAFAMLLLVTAVTAVRQVSRLRDRNPQAFWLMAISLAGLAGMIGHG 422
Db      354 ELLYWGVEGGLPEVWG---ILLAAQFCALRIYAAKRGR-----MAMLALFVPIITLHA 403
QY      423 LEDTVLYRPEASTLMLWLCIGAIASFQWPQ 451
Db      404 QLEYPFYH---SAIHWITF-IILITVWDQ 428

```

RESULT 8
S77086
hypothetical protein sll0737 - *Synechocystis* sp. (strain PCC 6803)
C;Species: *Synechocystis* sp.
A;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C;Accession: S77086
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,
O., K.; Okumura, S.; Shlimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasu
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocys*
s.
A;Reference number: S74322; MUID:97061201
A;Accession: S77086
A;Status: nucleic acid sequence not shown; translation not shown
A;molecule type: DNA
A;Residues: 1-861 <KAN>
A;Cross-references: EMBL:D64005; GB:AB001339; NID:g1001779; PIDN:BAA10778.1; PID:g1000
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C;Superfamily: *Synechocystis* hypothetical protein sll0737

```

Query Match      6.6%; Score 159; DB 2; Length 861;
Best Local Similarity 22.6%; Pred. No. 0.00073;
Matches 112; Conservative 63; Mismatches 145; Indels 176; Gaps 30;

QY  48 GFLAVVYGSAPFEVSS-ALGLG-----LAAIAAYW-----ALSLTDIDLR 88
    | | : : | | : : | | : | : | : | : | : | : | : | : | : | : |
Db  27 GLTALLYGLFTLLENGHSIMVGMWPWFVWQTWFWLTAIMCIMOIGQNKRLISL----- 80

QY  89 QATPIHMLVLLYMGWDALATGLSPVRA-----AAIVGLAKTLYILVFAALA 134
    || : : | : | | | | | | : | : | : | : | : | : | : | : | : |
Db  81 -GLGFDWLMAVAL-----AIAAMLSVGAEEFPQARWYGMIFCAFLAGLYGRSWL----- 129

QY  135 ARYLRRNP--RLRSL-----FSVVVITSLFVSYYGL-----NOMIYG 169
    : | | | : | | : : : : | : | : | : | : | : | : | : | : | : |
Db  130 -KVQTNPLVYSINKLLTFQGYVGFAFIVISLILMFTQTLLIPFWQSAATAREMGITKTFIFG 188

QY  170 VEEELATWDRNSVADFTSRVYSYLGNPMLIAIYLPPTAFSAATIGW-----KGLPKL 224
    : | | | | : | | : : : : | : | : | : | : | : | : | : | : | : |
Db  189 ILELQNWAP-----IGHQNVYAQYLV--LILPLLSYLIWLNEGKKRWFEWSV 232

QY  225 LAIAATGASSICLI---LTYSRGWLGFVAMIFWALLGLYWFQRPRLPAPWRRLPEPVVL 281
    | | | | | | | | | | | | : : : : : : : : | | | | | | | |
Db  233 -----ALSLGLIDFVYTTSSRGGLG-LATLLLLVLIIGIGLLR---QLPWRWML----- 276

QY  282 GGLVAVLLVAVLGL-EPLKVRVLSIFVGREDSN-----NEFINWVLAVLQMIQDR 331
    | | | : : : | : : | : | : | : | : | : | : | : | : | : | : |
Db  277 -GL-GTLAIAVVGIFICTNDRLLTSFTGIMGQAGQAFAYRLINEIG-W----RMGSAH 329

QY  332 PWLGIGPGNTAFNLVYPLY-----QQARETALSAYSVPLEVAVEGGLGLTAFAWLLLV 385
    | | | | | | | | | | : : : | : | : | : | : | : | : | : | : |
Db  330 PWTGIGLGNVP--LYQOLYRPWMAGRESEF--LYQLHSTPAQLFAELGV-----WGILI 379

QY  386 TAVTAVRQVSRLRRDRNPQ-----AFWLMASLAGLAGMLGHGLFDIVL 428
    : : : : : : | | : | : | : | : | : | : | : | : | : | : | : |
Db  380 PSLILIGLIWQILIRSLIKQNIIVNTKGNVDVSNVVKILIMTILTS-----ALLAYGVTSITLD 434

```


A:Experimental source: strain R1
C:Genetics:
A:Gene: DR1602
A:Map position: 1

Query Match 5.8%; Score 139.5; DB 2; length 388;
Best Local Similarity 28.4%; Pred. No. 0.0079;
Matches 84; Conservative 24; Mismatches 81; Indels 107; Gaps 20;

OY 20 SSFLHRLFGSLRAMRASSQLLVWSEALGFLLAVVYGSAPFVPSSALGLGLAIAIAYMAL 79
Db 146 ASFFAALLGQLRG-----RMQALAAALLAAYALFLHFTPHPG-GIGIVS----- 189
OY 80 LSLTDIDLROATPIHML---VLLYWGVDALATGLSPVRAALVGLAKLTLYLLVFALAAR 136
Db 190 -----ESANPYOYLNDALLSAGWGLRGL--LSVPTTALV-----LLGALAAR 229
OY 137 VL--RNPRLRSLFSV-VVITSLEVS-----VYGLNQWITYGVEELA 174
Db 230 PLQOKNPRAPLLLLGLGVVLTALGYCWAASGRLPFSKALWTPPYILYSAGLGLGI--LA 287
OY 175 TWVDRNSVADF-----TSRVSYLGNP--MLAAYLVPTTAESAAGVW-----RGWL 221
Db 288 CWV----VADSGWLPQKRLAPLTPGRNALAGYVLP-----ILIKWILLDQVQWMT 337
OY 222 PKLLAIAATGASSLCLILTYSR-----GGM--LGFVAMIFVWALLGLYWFQPR 267
Db 338 GRSQSIAS-----LLEMARSSFPGVGGMWYTLGYV--FAVW--LGLAMMAR 382

RESULT 12
E86038
Probable LPS biosynthesis enzyme waal [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: E86038
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Hiller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: E86038
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-402 <STO>
A:Cross-references: GB:AE005174; NID:g12518378; PIDN:AAG58769.1; GSPDB:GN00145; UWGP:250
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: waal

Query Match 5.7%; Score 138.5; DB 2; length 402;
Best Local Similarity 21.8%; Pred. No. 0.0096;
Matches 85; Conservative 67; Mismatches 151; Indels 87; Gaps 19;

OY 26 LFGSL--RAMRASSQLLVWSEALGFLLAVVYGSAPFVPSSALGLGLAIAIAY---MAL 79
Db 5 LFFSLEKKNW-----LAYWNRAL-VLFITYY-----FLGCTRYKHLIVIL 45
OY 80 LSLTDIDLROATPIHMLVL---LYWGDALATGL-----SPVRAALVGLAKLT-L 128
Db 46 MTTTIVLYLCKRPKHLYLSFKTFLFGSAVAILTAALLSLQSPDAGASMEIKAIIENT 105
OY 129 LVFALAARVLNRPRLRSLFSVVVITSLEFSVYGL-----NQWITYGVEELATWVD 178
Db 106 LICTTAIPITLRDEKREDEKIVFFS--FISALGLRCFSESLITYYKDYQOGIMPFADYRH 163
OY 179 RNSVADFTSRVSYLGNPNLLAAYLVPTTAESAAGVW--LGLAIAIATGASSLCL 238
Db 164 R-SISD--SMVFLF--PALLNLMLIKSAKYR-----ISFVLSVIFIFLI 203
OY 239 L-TYSRGGWLGFAVAMIFVWALLGLYWFQPRLPAPWRRLFPVVLGGLVAVLLVAVLGL 297

Db 204 LGTLRGAWLSVLVIGLIWILM-----FKQWKL-LLVGVMAIITALSVIFTHK 250
OY 298 LRVRLSTFEVGREDDSSNNFRINWLAVLQMIQDRPWLIGPGNTAFNLVYPL---YQQA 353
Db 251 EMTAKLTYKLOQTNSSRYRANGTQGSALDILLENPNYIGYGVNAVYKDYNNKRVIDYPEW 310
OY 354 RF-TALSAISVPLEVAVEGGLGLTAFAML 382
Db 311 TFRQSIGPHNEALFIFWGTGLGLVSLMML 340

RESULT 13
G75267
ABC transporter, permease protein, CysTW family - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: G75267
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J
S.; Shen, M.; Vamathavan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896
A:Accession: G75267
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-515 <WHI>
A:Cross-references: GB:AE002078; GB:AE000513; NID:g6460306; PIDN:AAF12029.1; PID:g646
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR2488
A:Map position: 1
C:Superfamily: sfub protein

Query Match 5.7%; Score 138; DB 2; length 515;
Best Local Similarity 23.6%; Pred. No. 0.013;
Matches 102; Conservative 51; Mismatches 154; Indels 126; Gaps 20;

OY 39 LLVWSEALGFL--LLAVVYGSAPFVPSSALG---LGLAIAIAYWALLSLTDIDLROATP 92
Db 129 LLIWGNLFENLPVMIRLAYGGEFARVPTNLIGAARTLGASGARAAM-----DYALPLALP 182
OY 93 ----IHWLVLLYWGVDALATGLSPVRAALVGLAKLTLYLLVFALAARVLNRPRLRSLF 148
Db 183 GLAAGAVLPLY--SALSFGI----PLALGGERYATLEVEIYTTALQRLPEASALIV 235
OY 149 SVVITSLEFSVYGLNQWITYGVEELATWVDRNSVADFTSRVSYLGNPNLLAAYLVPTTA 208
Db 236 GQLALTL-----VATWL-----YTR 250
OY 209 FSAAGVWGRWLPKLLAIAATGAS---SLCLIIYVSRGGWLGFAVAMIFVWALLG----- 260
Db 251 LTRGAGVATGGLPRARAGALAGITLLGTLTLYCFAP-----LLAVVWGLVGSAGPT 304
OY 261 -LYWFQPRLPAPWRMLF-----PVVLGGLVAVLLVAVLGLLEPLRVRLSIFVGREDDSN 314
Db 305 LLYW-----RGVNVDEQFPDLLGNTRLRFGALALLGATAL-----CG 340
OY 315 NFRINWLA---VLQMIQDRPWLIGPGNTA--FNLVYPLYQQAARTALSAVS-VPLEVA 368
Db 341 LYALGAWLARSRYLDLISLP-LMWSPVSLAVGYLLAAYVLAATLPMLIAAYTLLALPLL 399
OY 369 VEGGLGLTAFAMLLVTAVTAVKOVSRRLRDRNPQAFWMLASLAG-----LAGMLGHGL 423
Db 400 VRSLPALRAIPRLEFAARTL--GASPLAHRVTYTLPLPALRGGAALALATVIGRFG 457
OY 424 FDTVLYRPEASTL 436
Db 458 ATLVLTRPEWATL 470

RESULT 14
F83759
cytochrome d (bd-type) ubiquinol oxidase subunit I BH0878 (imported) - *Bacillus halodurans*
C/Species: *Bacillus halodurans*
C/Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
C/Accession: F83759
R/Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A/Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and
A/Reference number: A83650; MUID:20263314
A/Accession: F83759
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-443 <STO>
A/Cross-references: GB:AP001510; GB:BA000004; NID:g10173440; PIDN:BAB04597.1; GSPDB:GN00
A/Experimental source: strain C-125
C/Genetics:
A/Gene: BH0878
C/Superfamily: cytochrome d complex terminal oxidase chain I

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2002, 23:03:58 ; Search time 23.92 Seconds
(without alignments)
715.823 Million cell updates/sec

Title: US-09-887-038-3
Perfect score: 2426
Sequence: 1 MTVMQTLTFAHYQPQWGHG.....WQPPSKQLPPEAEHSDKM 467

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result NC.	Score	Query Match	Length	DB ID	Description
1	152	6.3	426	1	EXOO_RHIME
2	134.5	5.5	660	1	FHUB_ECOLI
3	122.5	5.0	404	1	RNAL_SALTY
4	122	5.0	425	1	NUON_ECOLI
5	121.5	5.0	333	1	YK26_PSEAE
6	120.5	5.0	815	1	AOX1_AERPE
7	119	4.9	463	1	NU4M_STRPU
8	118.5	4.9	461	1	Y608_HAEIN
9	118.5	4.9	535	1	LAT2_HUMAN
10	118.5	4.9	614	1	FE0B_SYNY3
11	118	4.9	647	1	CCMF_ECOLI
12	117	4.8	662	1	YANK_RHISN
13	116.5	4.8	317	1	EXOZ_RHIME
14	116.5	4.8	660	1	CCMF_PSEFL
15	116	4.8	533	1	LAT2_RAT
16	115.5	4.8	419	1	YHHS_ECOLI
17	115.5	4.8	439	1	POTE_ECOLI
18	115.5	4.8	547	1	YEOO_ECOLI
19	115	4.7	504	1	CAIT_ECOLI
20	115	4.7	504	1	PUPP_HAEIN
21	115	4.7	531	1	LAT2_MOUSE
22	115	4.7	533	1	MYIN_SYNY3
23	115	4.7	550	1	YFBI_ECOLI
24	114.5	4.7	414	1	MTR_ECOLI
25	114.5	4.7	3011	1	POLG_HCVI
26	114	4.7	531	1	NUON_MYCTU
27	113.5	4.7	431	1	NOOE_THETH
28	113.5	4.7	3011	1	POLG_HCVH
29	113	4.7	379	1	HCAT_ECOLI
30	112.5	4.6	306	1	YEDA_ECOLI
31	112	4.6	416	1	RHLD_GORGO
32	111	4.6	404	1	Y4XM_RHISN
33	111	4.6	405	1	PE21_MOUSE

34	111	4.6	434	1	YHFT_ECOLI	P45546	escherichia
35	111	4.6	441	1	Y037_MYCTU	P71607	mycobacteri
36	111	4.6	471	1	NORM_CAUCR	P58163	caulobacter
37	111	4.6	543	1	TIC2_CHLMU	Q9pjp6	chlamydia m
38	111	4.6	587	1	DSD2_PSEAE	Q91104	pseudomonas
39	110.5	4.6	447	1	NI4M_ANOGA	P34852	anopheles g
40	110.5	4.6	475	1	MMR_STRCO	P1545	streptomyce
41	110.5	4.6	563	1	YVFH_BACSU	P71067	bacillus su
42	109.5	4.5	442	1	YA17_SYNY3	P72935	synechocyst
43	109.5	4.5	459	1	PUCG_RHOSH	Q02443	rhodobacter
44	109.5	4.5	485	1	PTDA_ECOLI	P24241	escherichia
45	109	4.5	428	1	BCH2_RHOCA	P26171	rhodobacter

ALIGNMENTS

RESULT	1	EXOO_RHIME	STANDARD;	PRT;	426 AA.
ID	EXOO_RHIME				
AC	002729;				
DT	01-JUL-1993 (Rel. 26, Created)				
DT	01-JUL-1993 (Rel. 26, Last sequence update)				
DT	30-MAY-2000 (Rel. 39, Last annotation update)				
DE	EXOPOLYSACCHARIDE PRODUCTION PROTEIN EXOQ.				
GN	EXOQ.				
OS	Rhizobium meliloti (Sinorhizobium meliloti).				
OG	Plasmid pSymb (megaplasmid 2).				
OC	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;				
OC	Rhizobiaceae; Sinorhizobium.				
OX	NCBI_TaxID=382;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-RCR2011 / SU47;				
RX	MEDLINE-93177026; PubMed-8439670;				
RA	Mueller P., Keller M., Weng W.M., Quandt J., Arnold W., Puehler A.;				
RT	"Genetic analysis of the Rhizobium meliloti exoYfg operon: ExoY is				
RT	homologous to sugar transferases and ExoQ represents a transmembrane				
RT	protein.";				
RL	Mol. Plant Microbe Interact. 6:55-65(1993).				
CC	-I- FUNCTION: INVOLVED IN THE PRODUCTION OF EXOPOLYSACCHARIDE.				
CC	-I- PATHWAY: EXOPOLYSACCHARIDE BIOSYNTHESIS.				
CC	-I- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; L05588; AAA26266.1; -				
KM	Exopolysaccharide synthesis; Transmembrane; Plasmid.				
FT	TRANSMEM	11	31		POTENTIAL.
FT	TRANSMEM	35	55		POTENTIAL.
FT	TRANSMEM	65	85		POTENTIAL.
FT	TRANSMEM	117	137		POTENTIAL.
FT	TRANSMEM	156	176		POTENTIAL.
FT	TRANSMEM	178	198		POTENTIAL.
FT	TRANSMEM	203	223		POTENTIAL.
FT	TRANSMEM	230	250		POTENTIAL.
FT	TRANSMEM	325	345		POTENTIAL.
FT	TRANSMEM	361	381		POTENTIAL.
SO	SEQUENCE	426 AA;	46486 MW;		CRS0281B726F9195 CRC64;

Query Match 6.3%; Score 152; DB 1; Length 426;

Best Local Similarity 23.5%; Pred. No. 0.0058;

Matches 100; Conservative 62; Mismatches 146; Indels 118; Gaps 22;

OY 43 SEALGFLAAVYGSAPFVPSALGLAIAAYMAL-LSLTIDILRQ----- 89

```

Db 14 NEVYGFALAL-----SLFVFAVSARFGVSYLAVYGLWLPVLVDYRKVLGNVASYLMIF 69
QY 90 ATPIHMLVLLKQVD---ALATGLSPVR--AAALVGLAKLTLYLVFALARVLRNPLR 144
Db 70 APTIFACITIFWSAAPSLSLRIGIQLSHVVCALIAMRTIDIRLTRGMIAGA----- 122
QY 145 SLFSVVVITSLFVSYSYGLNOMIYGVETATWDRNSVADFTSRVYSYLG--NPMLAA 201
Db 123 ----ATVLLYSLEFGTYHY-----DPLDGYTSFVGAFASKNQLGF 158
QY 202 YLVPTTAFSAAI-----GVWRGWLPKLLAIAATGASSL--CLILTSRGGWLGIVAM 252
Db 159 YASLGIFYFAFAAVFVLGKGLMVG-----AAGGAGLLAAYCLLTSQSATSVLTTAAV 210
QY 253 IFVWALLGLYWFQPLRPAPWRRLF--PVVLGLVAVLLVAVLGLLEPLRVRLSIFVGRE 310
Db 211 IGL--CLGMRAITALRPAS-RKLLFIAASVLAGVAAMVAGVD---LILGAF--GK 261
QY 311 DSSNNFRINWLAVLQMIQDRPWLIGIPGNTAFNLVYPLYOARFT----- 356
Db 262 DSTLTGRTYLMQOGIEAKVSPLVGVG-----YQAYWVGQFSEAERLMEEFYIGSR 312
QY 357 -ALSAYSVPLEVAVEGGLGLTAFAMLLVTAFTA-VKQVSR-LRRDRNQAFWLMASLA 413
Db 313 AGTFHNTFIEAVETGLIGLI---LITWVLTAFEGQLKRLLEDREDPESMVLF---- 364
QY 414 GLAGML 419
Db 365 GVGALL 370

```

RESULT 2

```

FHUB_ECOLI STANDARD; PRT: 660 AA.
ID FHUB_ECOLI
AC. P06972; P77372;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE FERRICHRONE TRANSPORT SYSTEM PERMEASE PROTEIN FHUB (FERRICHRONE UPTAKE
DE PROTEIN FHUB).
GN FHUB OR B0153.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=87014116; PubMed=3020380;
RA Koester W., Braun V.;
RT "Iron hydroxamate transport of Escherichia coli: nucleotide sequence
RT of the fhuc gene and identification of the protein.";
RL Mol. Gen. Genet. 204:435-442(1986).
RN 12
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / W3110;
RX MEDLINE=94261430; PubMed=8202364;
RA Fujita N., Mori H., Yura T., Ishihama A.;
RT "Systematic sequencing of the Escherichia coli genome: analysis of
RT the 2.4-4.1 min (110,917-193,643 bp) region.";
RL Nucleic Acids Res. 22:1637-1639(1994).
RN 13
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirtpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN 14
RP SEQUENCE FROM N.A.
RA Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,

```

```

RA Davis K., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,
RA Lashkari D., Lew H., Lin D., Namath A., Oefner P., Roberts D.,
RA Davis R.W.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RN 15
RP SEQUENCE OF 1-23 FROM N.A.
RX MEDLINE=88038363; PubMed=2823072;
RA Burkhardt R., Braun V.;
RT "Nucleotide sequence of the fhuc and fhud genes involved in iron
RT (III) hydroxamate transport: domains in Fhuc homologous to
RT ATP-binding proteins.";
RL Mol. Gen. Genet. 209:49-55(1987).
CC -1- FUNCTION: THIS IS ONE OF THE PROTEINS INVOLVED IN THE HIGH-
CC AFFINITY TRANSPORT OF IRON(III)-FERRICHRONE INTO THE E.COLI
CC CELL.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
CC -1- SIMILARITY: WITH INTEGRAL MEMBRANE COMPONENTS OF OTHER BINDING-
CC PROTEIN-DEPENDENT TRANSPORT SYSTEMS. BELONGS TO THE FECCD
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X04319; CAA27852.1; -.
DR EMBL; X04319; CAA27853.1; ALT_INIT.
DR EMBL; D26562; CAB20273.1; -.
DR EMBL; AE000124; AAC73264.1; -.
DR EMBL; U70214; AAB08583.1; -.
DR EMBL; X05810; CAA29256.1; -.
DR PIR; S07318; S07318.
DR PIR; S45222; S45222.
DR Ecogene; EG10303; fhuc.
DR InterPro; IPR000522; Feccd.
DR Pfam; PF01032; Feccd_family; 2.
DR ProDom; PD001557; Feccd; 2.
KW Transport; Iron transport; Transmembrane; Inner membrane;
KW Complete proteome.
FT TRANSMEM 5 25 POTENTIAL.
FT TRANSMEM 63 83 POTENTIAL.
FT TRANSMEM 93 113 POTENTIAL.
FT TRANSMEM 118 138 POTENTIAL.
FT TRANSMEM 147 167 POTENTIAL.
FT TRANSMEM 197 217 POTENTIAL.
FT TRANSMEM 240 260 POTENTIAL.
FT TRANSMEM 277 297 POTENTIAL.
FT TRANSMEM 303 323 POTENTIAL.
FT TRANSMEM 348 368 POTENTIAL.
FT TRANSMEM 391 411 POTENTIAL.
FT TRANSMEM 424 444 POTENTIAL.
FT TRANSMEM 447 467 POTENTIAL.
FT TRANSMEM 479 499 POTENTIAL.
FT TRANSMEM 528 548 POTENTIAL.
FT TRANSMEM 567 587 POTENTIAL.
FT TRANSMEM 607 627 POTENTIAL.
FT TRANSMEM 635 655 POTENTIAL.
FT TRANSMEM 11 11 L->V (IN REF. 5).
FT TRANSMEM 11 11 T->I (IN REF. 1 AND 2).
FT TRANSMEM 311 311 MISSING (IN REF. 1 AND 2).
FT TRANSMEM 364 364 MISSING (IN REF. 1 AND 2).
FT TRANSMEM 660 660 AA; 70422 MW; 427313E7B421B0BC CRC64;
SO SEQUENCE

```

```

Query Match 5.5%; Score 134.5; DB 1; Length 660;
Best Local Similarity 22.0%; Pred. No. 0.11;
Matches 105; Conservative 54; Mismatches 141; Indels 177; Gaps 23;

```

```

QY 1 MTWQTLTFHAHYQPQGWGSSFLHRLFGSLRAWRASSQLVWSEALGFLAVVYGSAPF 60
Db 175 MFLWSTGTLTQ---TDWGG---VERL-----WPQLIGVVMTLTLL-----L 209

```

QY	61	VPSALG-----LGLAIAAYWALISTDIDLROATPIHMLVLLWGVDAIATGLSP	112
Db	210	RPLTIMGDDGVARNGLALSLARLMAISLAIV-----ISALL-----	247
QY	113	VRAALVGLAKLFTLLVLEFALARVLRNDRRLSLFFSVVITSLEFVSYGLNOMYGVEE	172
Db	248	VNAVGLIGFIGLFAPLLAKMLGARRL-PRL-MLASLIGALIIWLS-----DQ	293
QY	173	LATWVDRNSVADFTSRVYSYLGPNLLAAYLVPTTAFSAAGVWRGM-LPKLLAIATG	231
Db	294	IILMLTRVMEVSTGVTALIGAPLL-----WLPRLRSISAPD	333
QY	232	-----ASSLCILITVS-RGWLIGFAMIFWAL---LGLYWFQRLRPAPWRRLFPV	279
Db	334	MKVNDRAVAERQHYLAIFALAGVLLMAVVALSFGRDHGMWTWASGALLEDIEMWRMPR	393
QY	280	VLGLVAVLLVAVGLLEPLRVRLSTIFVGRDSSNNFRINVMWLAVLQMIQDRPW-----L	334
Db	394	IMALLFAGVMIAVAG-----CIIQLTGPNMASPEVL	425
QY	335	GIGPGNTAFENLVPLYQARFTALSAISVPLEVAEGGLGLTAFAML---LVTAVT-	389
Db	426	GISSG-AAFGVYMLF-----LYPGNAFGWLLPAGSLGAAYTL	462
QY	390	AVRQVSRLRDRNPQAFWLMASLAGLAGMLGHGLFDVLYRPEAS-----TIWWL	439
Db	463	LIIMIAAGRGGSFPHML-----LAGMALSTAFMTMLMLQASGDPRAQVYLTWI	512

```

RESULT 3
REFAL_SALT
ID REFAL_SALT STANDARD; PRT; 404 AA.
AC *P26471;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE O-ANTIGEN LIGASE.
GN *REFAL OR WAL OR REBT.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LT2;
RX MEDLINE=92041612; PubMed=1657881;
RA MacLachlan P.R., Kadam S.K., Sanderson K.E.;
RT *Cloning, characterization, and DNA sequence of the rfaL region for
RT lipopolysaccharide synthesis in Salmonella typhimurium LT2."
RL J. Bacteriol. 173:7151-7163(1991).
CC -1- FUNCTION: ADDS THE O-ANTIGEN ON THE GLUCOSE(II) GROUP OF LPS.
CC -1- PATHWAY: LIPOPOLYSACCHARIDE CORE BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: SHOWS ALMOST NO SIMILARITY TO E.COLI RFAI.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC
DR EMBL; M73826; AAA27206.1; .
DR PIR; B41317; B41317.
DR StyGene; SG10340; rfaL.
KW Lipopolysaccharide biosynthesis; Ligase; Transmembrane.
FT TRANSMEM 66 84 POTENTIAL.
FT TRANSMEM 184 203 POTENTIAL.
FT TRANSMEM 227 244 POTENTIAL.
SO SEQUENCE 404 AA; 46034 MW; 1B84F7B3C757D2AB CRC64;

```

	Query Match	5.0%;	Score 122.5;	DB 1;	Length 404;	
	Best Local Similarity	20.8%;	Pred. No. 0.39;			
	Matches	68;	Conservative	63;	Mismatches 123;	Indels 73; Gaps 16;
QY	129 LVFALAARVLRNPRRLRSLLFSVVVITSLFVSYYGLN---QWITYGVEELATWDRNSVADF	185				
	:: :: ::: ::: ::: ::: ::: :					
Db	105 LLYTLIPVLKDETKETVAKIVLFS--FLTSLGRCLAESILYIEDV-----NKGIIMPF	157				
QY	186 TSRVSYLGPNPLLAAYLVPTTAFTAASAAIGWV--RGWLPKLIAAATGASSLCILITYSR	243				
	: :: : : : : : : : : : : : : : : : :					
Db	158 ISVAHRHMSDS---MVFLFP-----ALINIMLFRRKAIKLVLSAIYLFFILGTLSR	208				
QY	244 GGWIGFVAMIEFWALLGLYWFOQRLEPAPWRMLFPVVLGGLVAVLLAVLGL-----	295				
	: : : : : : : : : : : : : : : : :					
Db	209 GAWLAVALIGVLMAILNRW-----KLIG--VGAILLAIIGALVITQHNN	251				
QY	296 --EPLRVRLSIFVGREDDSS----NNFRINWVLAVLOMTIQDRPVLGIGPGNTAFNLVPL	349				
	: : : : : : : : : : : : : : : : : :					
Db	252 KRPD---EHLLYKKLOQTDDSSRYRTNGTGWTAMI---LIQENPIKGYGYGNDVYDGYNK	304				
QY	350 ----YQOARF-TALSAYSVPLEVAVEGGILGLTAFAML---LLVTAVTAVRQVSRLRD	400				
	- : : : : : : : : : : : : : : : : :					
Db	305 RYVDYPTWTEKESIGPHNTILYIWFSAIGLGLASLVLYLGAIIRETASTLRKV---E	359				
QY	401 RNPOAFWLMAFLAGMLGHGLFDTV	427				
	: : : : : : : : : : : : : : : : : :					
Db	360 ISPDYNAHLLEFLSFVGFGYIVRNFEQV	386				

RESULT	4
ID	NUON_ECOLI
AC	P3608; P78281;
DT	01-FEB-1994 (Rel. 28, Created)
DT	01-NOV-1997 (Rel. 35, Last sequence update)
DT	20-AUG-2001 (Rel. 40, Last annotation update)
DE	NADH DEHYDROGENASE I CHAIN N (EC 1.6.5.3) (NADH-UBIQUINONE
DE	OXIDOREDUCTASE CHAIN N) (NUO14).
GN	NUON OR B2276 OR Z3534 OR ECS3160.
OS	Escherichia coli, and
OS	Escherichia coli O157:H7.
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC	Escherichia.
OX	NCBI_TaxID=562, 83334;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=K12 / AN387;
RX	MEDLINE=93389724; PubMed=7690854;
RA	Weidner U., Geler S., Plock A., Friedrich T., Lelf H., Weiss H.;
RT	"The gene locus of the proton-translocating NADH: ubiquinone
RT	oxidoreductase in Escherichia coli. Organization of the 14 genes and
RT	relationship between the derived proteins and subunits of
RT	mitochondrial complex I.";
RL	J. Mol. Biol. 233:109-122(1993).
RN	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN=K12 / MG1655;
RX	MEDLINE=97426617; PubMed=9278503;
RA	Blaettner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA	Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA	Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA	Mau B., Shao Y.;
RT	"The complete genome sequence of Escherichia coli K-12.";
RL	Science 277:1453-1474(1997).
RN	[3]
RP	SEQUENCE FROM N.A.
RC	STRAIN=K12;
RX	MEDLINE=97349980; PubMed=9205837;
RA	Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,
RA	Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsunashi N.,
RA	Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,
RA	Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasubram S.,

```

RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA Yamagata S., Horiuchi T.;
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT - K12 genome corresponding to 50.0-68.8 min on the linkage map and
RT analysis of its sequence features.";
RL DNA Res. 4:91-113(1997).
RN [4]
RN SEQUENCE FROM N.A.
RP STRAIN-O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE-21074935; PubMed-11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grolbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [5]
RN SEQUENCE FROM N.A.
RP STRAIN-O157:H7 / RIMD 0509952;
RX MEDLINE-21156231; PubMed-11258796;
RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
CC -1- FUNCTION: TRANSFER OF ELECTRONS FROM NADH TO THE RESPIRATORY
CC CHAIN. THE IMMEDIATE ELECTRON ACCEPTOR FOR THE ENZYME IS BELIEVED
CC TO BE UBIQUINONE. DOES COUPLE THE REDOX REACTION TO PROTON
CC TRANSLOCATION AND THUS CONSERVES THE REDOX ENERGY IN A PROTON
CC GRADIENT.
CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
CC -1- SUBUNIT: COMPOSED OF 13 DIFFERENT SUBUNITS. SUBUNITS NUOA, H, J,
CC K, L, M, N CONSTITUTE THE MEMBRANE SECTOR OF THE COMPLEX.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X68301; CAA48373.1; -.
DR EMBL: AE000317; AAC75336.1; -.
DR EMBL: D90858; CAB22029.1; -.
DR EMBL: D90859; BAA16104.1; -.
DR EMBL: AE005459; AAG57405.1; -.
DR EMBL: AP002561; BAB36583.1; -.
DR PIR: S38323; S38323.
DR EcoGene; EG12093; nuon.
DR InterPro; IPR001750; Oxidored_q1.
DR Pfam; PF00361; oxidored_q1.1.
KW Oxidoreductase; NAD; Ubiquinone; Transmembrane; Complete proteome.
FT TRANSMEM 11 31 POTENTIAL.
FT TRANSMEM 45 65 POTENTIAL.
FT TRANSMEM 67 87 POTENTIAL.
FT TRANSMEM 99 119 POTENTIAL.
FT TRANSMEM 143 163 POTENTIAL.
FT TRANSMEM 175 195 POTENTIAL.
FT TRANSMEM 211 231 POTENTIAL.
FT TRANSMEM 237 257 POTENTIAL.
FT TRANSMEM 266 286 POTENTIAL.
FT TRANSMEM 313 333 POTENTIAL.
FT TRANSMEM 336 356 POTENTIAL.
FT TRANSMEM 395 415 POTENTIAL.
FT CONFLICT 63 66 NHLA -> TIWR (IN REF. 1).
FT CONFLICT 140 141 EP -> DA (IN REF. 1).

```

```

FT CONFLICT 208 217 AIRVLAIA -> QVRGAGDYR (IN REF. 1).
SQ SEQUENCE 425 AA; 45650 MW; 9818DD4FFDE08E1F CRC64;

Query Match
Best Local Similarity 23.3%; Pred. No. 0.44;
Matches 82; Conservative 48; Mismatches 114; Indels 108; Gaps 19;

OY 27 FGLRWRWRASSQLLWSEALGGLL--AVVY---GSAPFV-----PSSALG 67
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 87 FRQKRSLEASIKYFTLSAASSFLFGMALVVAQSGDLSEFVALGKNLGDMNEPLLIG 146
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 68 LGIAIAIAYWALLSLDIDLROATPIH-WLVLYWGVDA-LATGLSPVRAALVGLA-KL 124
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 147 FGLMIVGLGFKL-----SLVPFHLMTPDVYOGAPAPVSTFLATASKIAIFGVWRL 197
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 125 TLVY-----LLVFALARVLRN-----PRLSLFESVVIT 154
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 198 FLYAPVGDSEAIRVLAITAFASITFGNLMALSGTNRKLLGYSSISHLGLVLAIALO 257
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 155 SLFVSVYGLNOMI-----YGVIELATWDRNSVADFTSRVSYLG--NPMLAA 201
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 258 TGENSMNAVGVYLAGYLESSLGAFGVSVLSMSPYRPDAD--SLFSYRGLEFNIHPILAA 314
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 202 YLVPTAFSAAAGVWRGWLPKLLAIAATGASSLCILITYSRGGMGLGVAMIFVALLGL 261
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 315 -VMTVMMLSLAGIPMTLIGFKFYVLAVGVQAHL-----WVL--VGAVVGSIAIGL 362
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 262 YWF-----QPRLPAPWRRLFPVVLGG--LVAVLLVAVLGLP 297
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 363 YYLRVAVSLYLHAPEQPRDAP-SNMQYSA--GGIVVILSALLVLVGVWP 411
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 5
YK26_PSEAE STANDARD; PRT; 333 AA.
ID YK26_PSEAE STANDARD; PRT; 333 AA.
AC P39879; Q91289;
DN 01-FEB-1995 (Rel. 31, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN PA2026.
GN PA2026.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAO1;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RA "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
RN [2]
RP SEQUENCE OF 281-333 FROM N.A.
RC STRAIN=PAO8;
RX MEDLINE=91194546; PubMed=1849605;
RA Perry A.C.F., Bhriain N.N., Brown N.L., Rouch D.A.;
RT "Molecular characterization of the gor gene encoding glutathione
RT reductase from Pseudomonas aeruginosa: determinants of substrate
RT specificity among pyridine nucleotide-disulphide oxidoreductases.";
RL Mol. Microbiol. 5:163-171(1991).
RL -1- SIMILARITY: TO E.COLI YFEH.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```



```

Db      110 YMLYLLSGVLVLAASF-----TESGAPNVGWTLAPLTARIYT---PGIGLD 153
QY      199 LAAYLVPTTAFSAAGVWRGWLPKLEIAATGASCLILTYSRGWLGFWAMIFVWAL 258
      111      |      |      |      |      |      |      |      |      |      |
Db      154 LAA-----LAIFLFSLSVTIGTINFLVLIAMRARGI-----GW--FKMPMEIWSI 197
QY      259 LGLYWFQPRRLPAPWRKWLFPVVLGGLVAVLLVAVLGE----- 296
      1      |      |      |      |      |      |      |
Db      198 L-----FTVILMLWAFPPPLMWGALLLLDRNLGTEFFFLNPAGCALLWDHLEWFEFGHP 249
QY      297 -----PLRVRVLSIFVGREDSNNFRINWMLAVLQWIDRPMVG--IGCGNTAF 343
      1      |      |      |      |      |      |      |      |      |
Db      250 EYVYILFPALGAMADVISTESGKPIYAKRYILTAFL-IATIISEVVMHMHMETGTNIYT 308
QY      344 NLVYPL-----YQQA-----RFTALSAYSVPLEVAEGGLGLTAFAMLLVYAV 388
      1      |      |      |      |      |      |      |      |      |
Db      309 RLEYSTITLISITFEMAVMSFTFLTKGRLLVYTVPMLEAV-GALLN-----FIIGGS 360
QY      389 TAVRQVSRLLRDRNPQ-AFWLMAASLAG-LAGMLGHGLFDIVLYRPEASTLWV 438
      1      |      |      |      |      |      |      |      |      |
Db      361 TGV-YLGSIAIDGFRGTYYWVAHFHYILVGTVTGLIAGLY-----WV 404

```

```

RESULT 7
NU4M_STRPU STANDARD: PRT; 463 AA.
ID NU4M_STRPU
AC P15551;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 4 (EC 1.6.5.3).
GN NDA.
OS Strongylocentrotus purpuratus (Purple sea urchin).
OG Mitochondrion.
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoida; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC Strongylocentrotus.
OX NCBI_TaxID=7668;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89011951; PubMed=3172215;
RA Jacobs H.T., Elliott D.J., Math V.B., Farguharson A.;
RT "Nucleotide sequence and gene organization of sea urchin
RT mitochondrial DNA.";
RL J. Mol. Biol. 202:185-217(1988).
RC -1 - CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X12631; CAA31159.2; -.
CC PIR: S01508; S01508.
DR InterPro: IPR003918; NADHUB_oxdrdctse4.
DR InterPro: IPR001750; Oxidored_q1.
DR InterPro: IPR000260; Oxidored_q5_N.
DR Pfam: PF00361; oxidored_q1.1.
DR Pfam: PF01059; oxidored_q5_N; 1.
DR Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
SQ SEQUENCE 463 AA; 50507 MW; 5C8E73FE7D5CAAF0 CRC64;

Query Match 4.9%; Score 119; DB 1; Length 463;
Best Local Similarity 18.8%; Pred. No. 0.73;
Matches 90; Conservative 52; Mismatches 132; Indels 204; Gaps 17;

Oy 61 VPSAALGGLAIAIAYWALLSLTDIDLRQATPIHW-----LVLLY 100
Db 19 IPSNKLWAGAI FOSALLSLSLVNLN-----NHWTASWHNTLSILASDTISAPLIILSC 72

```

[illegible]

```

RESULT      8
Y608_HAEIN  STANDARD;          PRT;      461 AA.
ID Y608_HAEIN
AC 057486;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN HI0608.
GN HI0608.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OC NCBI_taxid=727;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL Science 269:496-512(1995).
CC -|- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -|- SIMILARITY: BELONGS TO THE NADC/P/PHO87 FAMILY OF TRANSPORTERS.
CC NADC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U32743; AAC22267.1; -
CC TIGR: HI0608; -

```

DR InterPro; IPR001898; Na_sulph_symp.
 DR Pfam; PF00939; Na_sulph_symp; 3.
 DR PROSITE; PS01271; NA_SULFATE; 1.
 KW Hypothetical protein; Transmembrane; Transport; Complete proteome.
 FT TRANSMEM 13 33 POTENTIAL.
 FT TRANSMEM 54 74 POTENTIAL.
 FT TRANSMEM 81 101 POTENTIAL.
 FT TRANSMEM 120 140 POTENTIAL.
 FT TRANSMEM 170 190 POTENTIAL.
 FT TRANSMEM 211 231 POTENTIAL.
 FT TRANSMEM 256 276 POTENTIAL.
 FT TRANSMEM 286 306 POTENTIAL.
 FT TRANSMEM 314 334 POTENTIAL.
 FT TRANSMEM 349 369 POTENTIAL.
 FT TRANSMEM 377 397 POTENTIAL.
 FT TRANSMEM 399 419 POTENTIAL.
 FT TRANSMEM 439 459 POTENTIAL.
 SQ SEQUENCE 461 AA; 49761 MW; B5E6F6965B38EF06 CRC64;

Query Match 4.9%; Score 118.5; DB 1; Length 461;
 Best Local Similarity 23.2%; Pred. No. 0.78;
 Matches 110; Conservative 64; Mismatches 165; Indels 135; Gaps 27;

QY 50 LLAVVYGSAPVPSSALGLGLAA-IAAVNALSLTDIDLQATPIHMLVLLYGVDAIAT 108
 DB 22 LEFVLLNLVLPFEKANGSLALIAFLAVLW-----LSEALHVTITALL---VPLAV 69
 QY 109 GLSPVRA-AAVLGLAKLTLYLV--FALAA-----RVLRNPRL---RSLIFSVMVIT 154
 DB 70 ALGLVSTKQALVGFADPTIFLFFGGFSLATALHIQKDKLIANKINALRGNLFIAVIY- 128
 QY 155 SLFVSVYGLNOMIYVEELA-----TWVDRNSVADETSRYISYLGPNLLAAYLV 204
 DB 129 -LFLITAFLSMWSNTATTAAMMLPLAMGLISQLDREK--DHNTYFVLLG----- 175
 QY 205 PTTAFSA--AIGVWRGWLPKLLAIATGASSLCILITYSRGWLGVAMIFVWAL-IGL 261
 DB 176 --IAYSASIGMGTLVSPPN--AIYASN-----LNLTFSDWLWGLPIMILLPMIGI 226
 QY 262 YW--FQPRLPAPWRK-----LEPV-----VLGIVAVLLVAVIGLE--- 296
 DB 227 LYIFKPKLHLNFEQTFENIEMNPMRILFIIFPVIALTWIFSGKINPISGLGLOKNI 286
 QY 297 ---PLRVVUSIFVGREDSNNFRINWLAVALQMODRPWGLI---GPGNTAFNLVYPL 349
 DB 287 ASFDSIVALLAAIV-----ICSTGVASWKQIQSNTDMGVLMLEFGGLT----LSAV 333
 QY 350 YQOARFTALSAVSPLEVAVEG---LIGLTAFAWLLVAVTAARQVSRRLRRDRNPQAF 406
 DB 334 LKDSGASKTLADSIIV--MIDQHFYLLIGLLVAAFIIFLTEFTS-----NTASNA 381
 QY 407 WLMASLAGLAGMLGHGFDVLYRPEASTLWMLCIGAIASEFWQPOPSKOLPPEA 460
 DB 382 LLVPIFISIAQSLG-----MEIGLALITIGIGASCAPMLPVAT--PPNA 423

RESULT 9
 LAT2_HUMAN STANDARD; PRT; 535 AA.
 AC Q9UHT5; Q9Y445; Q9UKQ6; Q9UKQ7; Q9UKQ8;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE LARGE NEUTRAL AMINO ACIDS TRANSPORTER SMALL SUBUNIT 2 (L-TYPE AMINO ACID TRANSPORTER 2).
 GN SLC7A8 OR LAT2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.

RX MEDLINE=20044753; PubMed=10574970;
 RA Rossier G., Meier C., Bauch C., Summa V., Sordat B., Verrey F., Kuhn L.C.;
 RA "LAT2, a new basolateral 4F2hc/CD98-associated amino acid transporter of kidney and intestine.";
 RT J. Biol. Chem. 274:34948-34954(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=99178268; PubMed=10080183;
 RA Borsani G., Bassi M.T., Sperandio M.P., De Grandi A., Buoninconti A., Riboni M., Manzoni M., Incerti B., Pepe A., Andria G., Ballabio A., Sebastio G.;
 RT "SLC7A7, encoding a putative permease-related protein, is mutated in patients with lysinuric protein intolerance.";
 RL Nat. Genet. 21:297-301(1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99321901; PubMed=10391915;
 RA Pineda M., Fernandez E., Torrents D., Estevez R., Lopez C., Camps M., Lloberas J., Zorzano A., Palacin M.;
 RT "Identification of a membrane protein, LAT-2, that co-expresses with 4F2 heavy chain, an L-type amino acid transport activity with broad specificity for small and large zwitterionic amino acids.";
 RL J. Biol. Chem. 274:19738-19744(1999).
 CC -!- FUNCTION: SODIUM-INDEPENDENT, HIGH-AFFINITY TRANSPORT OF LARGE NEUTRAL AMINO ACIDS. HAS HIGHER AFFINITY FOR L-PHENYLALANINE THAT LAT1. L-ALANINE IS TRANSPORTED AT PHYSIOLOGICAL CONCENTRATIONS. PLAYS A ROLE IN BASOLATERAL (RE) ABSORPTION OF NEUTRAL AMINO ACIDS.
 CC -!- SUBUNIT: DISULFIDE-LINKED HETERODIMER WITH THE AMINO ACID TRANSPORT PROTEIN SLC3A2/4F2HC.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
 CC -!- TISSUE SPECIFICITY: MAINLY EXPRESSED IN KIDNEY AND INTESTINE.
 CC -!- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MANY AMINO ACID PERMEASES. GLYCOPROTEIN-ASSOCIATED AMINO ACID TRANSPORTERS (GPAT) SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF171669; AAF20381.1; -.
 DR EMBL; Y18483; CAB40137.1; -.
 DR EMBL; AF135830; AAF05697.1; -.
 DR EMBL; AF135829; AAF05696.1; -.
 DR EMBL; AF135828; AAF05695.1; -.
 DR MIM; 604235; -.
 DR InterPro; IPR002293; AA_rel-permease_1.
 DR InterPro; IPR002027; Amino_acid_permease.
 DR Pfam; PF00324; aa_permeases; 1.
 KW Transport; Amino-acid transport; Transmembrane.
 FT TRANSMEM 40 60 POTENTIAL.
 FT TRANSMEM 72 92 POTENTIAL.
 FT TRANSMEM 113 133 POTENTIAL.
 FT TRANSMEM 155 175 POTENTIAL.
 FT TRANSMEM 189 209 POTENTIAL.
 FT TRANSMEM 231 251 POTENTIAL.
 FT TRANSMEM 268 288 POTENTIAL.
 FT TRANSMEM 310 330 POTENTIAL.
 FT TRANSMEM 362 382 POTENTIAL.
 FT TRANSMEM 388 408 POTENTIAL.
 FT TRANSMEM 424 444 POTENTIAL.
 FT TRANSMEM 447 467 POTENTIAL.
 FT TRANSMEM 225 225
 FT CONFLICT 401 401 V -> G (IN REF. 2).
 FT CONFLICT 503 503 G -> R (IN REF. 3).
 SQ SEQUENCE 535 AA; 58381 MW; AC129146353F1E47 CRC64;

Query Match 4.9%; Score 118.5; DB 1; Length 535;
Best Local Similarity 22.6%; Pred. No. 0.89;
Matches 89; Conservative 56; Mismatches 123; Indels 125; Gaps 19;

QY 1 MTVMQTLTFAHYQ--POQWGHSSFLHRLFGS---LRWRASSQLLVWSEAL-----GC 48
DB 85 ITVVGALCYAELGVTIPKSGDYSVKDIQGLAGFLRLTAIVLVYIPTNOAVIALFPESN 144
QY 49 FLAAVYGSAPFVPSALGLGIAIAAYWALLSLTDID--LRQATPIHML----- 96
DB 145 YVLQPLEPPTC-FPPESGLRL-LAAL-----CLLLLTWVNCSSVWMATRVQDIFTAGKLAL 198
QY 97 -VLLVWGVDALATG---LSPVRA-----AALVGLA----- 122
DB 199 ALIIMGIQICKGEYFWLEPKNAFENFOEPDGLVALAFLOCSFAYCGWNLNYTEEL 258
QY 123 -----KLTLYLLVFALAARV-LRNPRLSLFSVVVITSLFVSVYGLNOM 166
DB 259 VDPYKMLPRAIFISIPLVTFVYVFANAVYTA MSP--QELLASNAVAVTFEGEKLGYMAW 316
QY 167 IYGV-EIATWVDNRNSVADFTSRVSYLGNPNLLAAYLVPTTAESAAGVWRGWLPKL 225
DB 317 IMPISVALSTFGVNGSLFTSRLE-----FAGAREGHLPSTVL 354
QY 226 -----AIAATGASSLCILIT---YSRGWLGFAVAMIF---VMALLGYWFOF 266
DB 355 AMIHVKRCTPIPALFTCISTLLMLVTSMDYTLINYYGVFINYLFGVTVAGQIVLRMKP 414
QY 267 RLPAFWR-RWLEFPVVLGGLVAVLVAVLGLEPL 298
DB 415 DIPRPKINLLEPIIYLLFWAFLLVFLVSLMSEPV 447

RESULT 10
FEOB_SYNY3
ID FEOB_SYNY3 STANDARD; PRT; 614 AA.
AC P73182;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE FERROUS IRON TRANSPORT PROTEIN B HOMOLOG.
GN FEOB OR SLR1392.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxId=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugitara M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K.,
RA Okumura S., Shimo S., Takeuchi C., Wada T., Watanabe A.,
RA Yamada M., Yasuda M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -1- FUNCTION: PROBABLE GTP-DRIVEN TRANSPORTER OF FERROUS ION (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D90904; BAAL7208.1; -.
DR InterPro; IPR003373; Feob.
DR Pfam; PF02421; Feob; 1.
KW Iron transport; Transport; Transmembrane; GTP-binding;

KW Complete proteome.
FT NP_BIND 25 32 GTP (POTENTIAL).
FT NP_BIND 71 75 GTP (POTENTIAL).
FT TRANSMEM 87 107 POTENTIAL.
FT TRANSMEM 113 133 POTENTIAL.
FT TRANSMEM 234 254 POTENTIAL.
FT TRANSMEM 273 293 POTENTIAL.
FT TRANSMEM 297 317 POTENTIAL.
FT TRANSMEM 344 364 POTENTIAL.
FT TRANSMEM 374 394 POTENTIAL.
FT TRANSMEM 404 424 POTENTIAL.
FT TRANSMEM 466 486 POTENTIAL.
FT TRANSMEM 504 524 POTENTIAL.
FT TRANSMEM 558 578 POTENTIAL.
FT TRANSMEM 585 605 POTENTIAL.
SQ SEQUENCE 614 AA; 67405 MW; 24587CF8F486713A CRC64;

Query Match 4.9%; Score 118.5; DB 1; Length 614;
Best Local Similarity 20.5%; Pred. No. 1;
Matches 83; Conservative 66; Mismatches 158; Indels 97; Gaps 17;

QY 17 WGHSSFLHRLFGSLRWARRASSQLLVWSEALGFLLAVY-GSA--PVPSSAL----- 66
DB 268 WQSNILEPLFSPL-----PTIQLGLLDGIMTGFALLSEVPILVAIFETVMG 315
QY 67 ---GLGLAIAAYW--ALLSLTDIDLQATPIHMLVLLVWGVDALATGSPVRAALVGL 121
DB 316 ILEGSGYLSRAAYLMDALMGRGLDGRS-----FVLQMGFGCNVPAINGTNRVMSRGM 369
QY 122 AKLTLYLLVFALAARVLRNPRLSLFSVVV-----TSLFVSVYGLNOMIYVEELA 174
DB 370 RLISMIVIFSLC-----SARLQVEVFILAAVMPGTQGAIALFL-----LY----LM 412
QY 175 TWVDRNSVADFTSRVSYLGNPNL--LAAYLVPTTAESAAGVW---KGWLPKLLAIA 228
DB 413 SFVAFTVAAILSRFHYQARDPVELLPYRLPT--FKQVFLRVWGENREFFVARLSMF 470
QY 229 ATGASSLCILITYSRGWLGFAVAMIFWALLGLVWFQPRLPAPWRRLFPVVLGGLVAVL 288
DB 471 VIGSSLIWFLTSFPGS-----TGLETFAGRIGSVFQPLMNPGLINPFLTIS 517
QY 289 LVAVLGLEPLRVRLSTFVGREDSSNNFRINWLAVLQMIQDRPWLIGDPGNTAFNLVY- 347
DB 518 LIFGVAKVQIAALTYIYGLNNS-----AVSDQIHSTVTFAGGFSYCLFSLIYI 568
QY 348 -----PLYQARFTALSAYSV--PLEVAVEGCHLGLTAPAWL 382
DB 569 PCLTTIGAIWGESKSLAYTAISVATPLVTAWLFSFIYQSFWSL 612

RESULT 11
CCMF_ECOLI
ID CCMF_ECOLI STANDARD; PRT; 647 AA.
AC P33927;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CYTOCHROME C-TYPE BIOGENESIS PROTEIN CCMF.
GN CCMF OR B2196.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxId=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / BHB2600;
RA Richterich P., Lakey N., Gryan G., Jaehn L., Mintz L., Robison K.,
RA Church G.M.;
RL Submitted (OCT-1993) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;

```

RX MEDLINE=97426617; PubMed=9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL science 277:1453-1474(1997).
[3]
RN CHARACTERIZATION, AND GENE NAME.
RP MEDLINE=95362656; PubMed=7635817;
RX Thoeny-Meyer L., Fischer F., Kunzler P., Ritz D., Hennecke H.;
RT "Escherichia coli genes required for cytochrome c maturation.";
RL J. Bacteriol. 177:4321-4326(1995).
CC -1- FUNCTION: REQUIRED FOR THE BIOGENESIS OF C-TYPE CYTOCHROMES.
CC POSSIBLE SUBUNIT OF A HEME LYASE.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE CCMF/CYCK/CCL1/NRFE/CCSA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U00008; AAA16388.1; -.
DR EMBL; AE000309; AAC75256.1; -.
DR EcoGene; EG12054; ccmf.
DR InterPro; IPR002541; CytC_asm.
DR InterPro; IPR003567; CytC_bilog.
DR Pfam; PF01578; CytC_asm; 1.
KW Cytochrome c-type biogenesis; Transmembrane; Inner membrane;
KW Complete proteome.
FT TRANSMEM 9 29 POTENTIAL.
FT TRANSMEM 42 62 POTENTIAL.
FT TRANSMEM 94 114 POTENTIAL.
FT TRANSMEM 121 141 POTENTIAL.
FT TRANSMEM 176 196 POTENTIAL.
FT TRANSMEM 211 231 POTENTIAL.
FT TRANSMEM 236 256 POTENTIAL.
FT TRANSMEM 274 294 POTENTIAL.
FT TRANSMEM 312 332 POTENTIAL.
FT TRANSMEM 351 371 POTENTIAL.
FT TRANSMEM 390 410 POTENTIAL.
FT TRANSMEM 425 445 POTENTIAL.
FT TRANSMEM 448 468 POTENTIAL.
FT TRANSMEM 482 502 POTENTIAL.
FT TRANSMEM 609 629 POTENTIAL.
SQ SEQUENCE 647 AA; 71389 MW; 162316A1ED156DA8 CRC64;

Query Match 4.9%; Score 118; DB 1; length 647;
Best Local Similarity 21.4%; Pred. No. 1.1;
Matches 112; Conservative 75; Mismatches 158; Indels 178; Gaps 30;

QY 23 LHRLEGLSLRA--WRASSQLLVW---SEALGGFLAVVYGSAPFVPSSALGLGLAA----- 72
   ::::| | |::| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 22 VPLMGVARGDARMAASSRLEFAWLLFMSVAGAFVLV---NAFVNDEFTVTVYASNSNT 77
   ::::| | |::| | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 73 -----IAAYWALLSTDLIDLRQATPIHMLVLLYWGVDALATGLSPV-----RAALV 119
   ::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 78 QLPVWYRVAATWG-----AHESGLLLVLLMSGWTFAVAIFSQRTPLDIVARVLAIM 129
   ::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 120 GIARKLLTYLIV-----PALAARVLRNPRLR--SLFSVWVITSLFVSVYGLN 164
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 130 GWVSYVGFLLFILFTSNPFSRTLPNEPIEGRDL-NPLLDPGLLFHPPL--LYMGVYGF 185
   : : | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 165 -QWIVGEELATWVDRNSVADFTSRVYSYLGPNLLAAYLVPT-----TAFSAAIGVW 217
   : : | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 186 VAFAFALASLLSGLRDLSTYARPT-----RPWTLAAMFTLLGLIVLGSAMAYVELG-W 236
   : : | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

QY      218 RGW-----LPKLIATA-----ATGASSCLILTY-SRG 244
          ||              :| | : |         | | | | : | : |
Db      237 GGWWEVDVENASFMPLVGTALMHSIAVTEQRASFKAWTLLLAISAFSLCLGTFLVRS 296
          ||              :| | : |         | | | | : | : |
QY      245 GWL-----GFVAMIEFWALLG--LYWFQPR-----LPAPWRR--WLFPV 279
          | |              :| | : |         | | | | : | : |
Db      297 GVLVSVAHAFASDPARGMFILAFMWLVIGSLLLFAARGHKVRSRVNNAIAMSRESLLANN 356
          ||              :| | : |         | | | | : | : |
QY      280 VLGLAVALLVAVLG-LEPLRVRL---SIFVGREDDSSNNFRINWVLAVLOMIQDRPWLG 335
          || || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      357 VL-LYAAMLVLLGTLPLVHKQLGLGISISIGEPPFNTMF---TWL----MVPEALLIG 407
          ||              :| | : |         | | | | : | : |
QY      336 IGP-----GNTAENLVYPLOYARFTALSAYSVPLEVAVEGGL 374
          :||              :| | : |         | | | | : | : |
Db      408 VGPLVRMGGRDPRPKIRNLLIAFISTLVLSLLPMLFEKVVAWT-----VL 454
          ||              :| | : |         | | | | : | : |
QY      375 GLTAFAWLLLVTAVTAVRQVSRRLRDNRNPQAFW-LMASIAGLA 416
          ||              :| | : |         | | | | : | : |
Db      455 GLAMACMAVLAVLIAEALRLIS-RGITKTFSYWGMAHLGLA 495
          ||              :| | : |         | | | | : | : |

RESULT 12
YANK_RHISN
ID      YANK_RHISN        STANDARD;       PRT;       662 AA.
AC      P55583;
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      01-NOV-1997 (Rel. 35, Last annotation update)
DE      HYPOTHETICAL 72.8 KDA PROTEIN YANK.
GN      YANK.
OS      Rhizobium sp. (strain NGR234).
OC      Plasmid sym pNGR234a.
CC      Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC      Rhizobiaceae; Rhizobium.
OX      NCBI_TaxID=394;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=97305956; PubMed=9163424;
RA      Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
RA      Perret X.;
RT      "Molecular basis of symbiosis between Rhizobium and legumes.";
RL      Nature 387:394-401(1997).
CC      -|- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC      -|- SIMILARITY: NONE OBVIOUS.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outpost -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
-----
CC
DR      EMBL; AE000087; AAB91790.1; -
KW      Hypothetical protein; Plasmid; Transmembrane.
FT      TRANSMEM 10 30 POTENTIAL.
FT      TRANSMEM 46 66 POTENTIAL.
FT      TRANSMEM 68 88 POTENTIAL.
FT      TRANSMEM 101 121 POTENTIAL.
FT      TRANSMEM 167 187 POTENTIAL.
FT      TRANSMEM 193 213 POTENTIAL.
FT      TRANSMEM 217 237 POTENTIAL.
FT      TRANSMEM 263 283 POTENTIAL.
FT      TRANSMEM 285 305 POTENTIAL.
FT      TRANSMEM 312 332 POTENTIAL.
FT      TRANSMEM 342 362 POTENTIAL.
FT      TRANSMEM 373 393 POTENTIAL.
FT      TRANSMEM 394 414 POTENTIAL.
FT      TRANSMEM 432 452 POTENTIAL.
FT      TRANSMEM 460 480 POTENTIAL.
FT      TRANSMEM 485 505 POTENTIAL.
SQ      SEQUENCE 662 AA; 72842 MW; 86465166D141B25D CRC64;
```



```
RC STRAIN=09906;
RA Yang C.H., Azad H.R., Cooksey D.A.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: REQUIRED FOR THE BIOGENESIS OF C-TYPE CYTOCHROMES.
CC POSSIBLE SUBUNIT OF A HEME LYASE (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE CCMF/CYCK/CCL1/NRFE/CCSA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U44827; AAC44226.1; -.
DR InterPro: IPR002541; Cytc_asm.
DR InterPro: IPR003567; Cytc_blog.
DR Pfam: PF01578; Cytc_asm; 1.
KW Cytochrome c-type biogenesis; Transmembrane; Inner membrane.
FT TRANSMEM 15 35 POTENTIAL.
FT TRANSMEM 49 69 POTENTIAL.
FT TRANSMEM 99 119 POTENTIAL.
FT TRANSMEM 126 146 POTENTIAL.
FT TRANSMEM 181 201 POTENTIAL.
FT TRANSMEM 216 236 POTENTIAL.
FT TRANSMEM 250 270 POTENTIAL.
FT TRANSMEM 279 299 POTENTIAL.
FT TRANSMEM 318 338 POTENTIAL.
FT TRANSMEM 360 380 POTENTIAL.
FT TRANSMEM 395 415 POTENTIAL.
FT TRANSMEM 426 446 POTENTIAL.
FT TRANSMEM 450 470 POTENTIAL.
FT TRANSMEM 493 513 POTENTIAL.
FT TRANSMEM 618 638 POTENTIAL.
SQ . SEQUENCE 660 AA; 72172 MW; 3C6F4B578C56436B CRC64;

Query Match 4.8%; Score 116.5; DB 1; Length 660;
Best Local Similarity 20.3%; Pred. No. 1.4;
Matches 107; Conservative 58; Mismatches 180; Indels 181; Gaps 21;

OY 3 VWQTLTFAHYOPQMGHSSFLHRLFGSL-----RAW 33
DB 39 LMSLSA----QPARMQGFAFLFAFGCLTYAFMTDDFSVAVYANNSNSALPMYYLFSAVW 94
OY 34 RA-SSQLVWSEALG-----GFLAVVYGSAPF-- 60
DB 95 GAHEGSLMLWALILGGWTFAVSVFSRQLPQVMLARVLAVMGIMISIGFLFLIMTSNPFSSR 154
OY 61 ----VPSSALGLGLAIAIAYWALLSTLTDIDLRQATPIHMLVLLYWGVDALATGLSPVRAA 116
DB 155 MLPQIPADGHDNLPL-----LDIGLIVHPP---MLYMG----- 185
OY 117 ALVGLAKLTLLVLFALARVLRNPLR-----SLFSVVITSLFVSVYGLNQIYGV 170
DB 186 -YVG-----FSVAFAFAIAALLGRLDAAMARMSRPWTIVAWAFILGIGITLGSWAYYE 238
OY 171 ELATWVDNSVADFTSRVSYIIGNPILAAVLVPTTAFSAAI---GWRGWLPKLLA 226
DB 239 LGWGWFMFMDPVENAS-----FMDPLVGTALIHSLAVTEKRGVFKSW-TVLLA 285
OY 227 IATGASSICLLILYSRGGMIGFVAMIFVWALLGLYWEQRLPAPWRMLFPVVLGLVA 286
DB 286 IAAFSLTLAAFLVRS-----GVLTSVHAF-----SDPER-----GVFTI. 321
OY 287 VLLVAVLGLLEPLRVRVLSIFVGREDDSSNNFRINWVLAVLQMIQDRPWLIGPGNTAFNLV 346
DB 322 IFLLFVVG-----GSLTLFVRAPVVKSGVGNLMSRETLLGLNNLVLVVAASMTLLGTL 375
OY 347 YPLYQGARFTALSAISVPLEVAVEGGLGLTAFAWLLLVTAVTAYROVSRLRRDRNPQAF 406
```

```
DB 376 YPLVLDALSGAKLSVGPPIYENALFIPLMGL-----LMVMAVGV-----LYRWKDPVK 424
OY 407 WLM-----ASLAGIAGMLGHGLEDTVLYRPEASTLWMLCIG 442
DB 425 WLVGMLAPVLLGSVALAVVAG-IAYGFNMAVLATPELLAAMVLAG 469

RESULT 15
LAT2_RAT
ID LAT2_RAT STANDARD; PRT; 533 AA.
AC Q9WVR6;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, last sequence update)
DT 20-AUG-2001 (Rel. 40, last annotation update)
DE LARGE NEUTRAL AMINO ACIDS TRANSPORTER SMALL SUBUNIT 2 (L-TYPE AMINO
DE ACID TRANSPORTER 2).
GN SLC7A8 OR LAT2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Small intestine;
RX MEDLINE=99321902; PubMed=10391916;
RA Segawa H., Fukasawa Y., Miyamoto K., Takeda E., Endou H., Kanai Y.;
RT "Identification and functional characterization of a Na+-independent
RT neutral amino acid transporter with broad substrate selectivity.";
RL J. Biol. Chem. 274:19745-19751(1999).
CC -1- FUNCTION: SODIUM-INDEPENDENT, HIGH-AFFINITY TRANSPORT OF LARGE
CC NEUTRAL AMINO ACIDS. HAS HIGHER AFFINITY FOR L-PHENYLALANINE THAT
CC LAT1. L-ALANINE IS TRANSPORTED AT PHYSIOLOGICAL CONCENTRATIONS.
CC PLAYS A ROLE IN BASOLATERAL (RE)ABSORPTION OF NEUTRAL AMINO ACIDS.
CC -1- SUBUNIT: DISULFIDE-LINKED HETERODIMER WITH THE AMINO ACID
CC TRANSPORT PROTEIN SLC3A2/4F2HC.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MANY AMINO ACID
CC PERMEASES. GLYCOPROTEIN-ASSOCIATED AMINO ACID TRANSPORTERS (GPRAT)
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB024400; BAA82517.1; -.
DR InterPro: IPR002293; AA_rel_permease_1.
DR InterPro: IPR002027; Amino_acid_permease.
DR Pfam: PF00324; aa_permeases; 1.
KW Transport; Amino-acid transport; Transmembrane.
FT TRANSMEM 41 61 POTENTIAL.
FT TRANSMEM 73 93 POTENTIAL.
FT TRANSMEM 114 134 POTENTIAL.
FT TRANSMEM 156 176 POTENTIAL.
FT TRANSMEM 190 210 POTENTIAL.
FT TRANSMEM 232 252 POTENTIAL.
FT TRANSMEM 269 289 POTENTIAL.
FT TRANSMEM 311 331 POTENTIAL.
FT TRANSMEM 363 383 POTENTIAL.
FT TRANSMEM 389 409 POTENTIAL.
FT TRANSMEM 423 443 POTENTIAL.
FT TRANSMEM 448 468 POTENTIAL.
SQ . SEQUENCE 533 AA; 58190 MW; 99479DB60DA69DF0 CRC64;
```

```
Query Match 4.8%; Score 116; DB 1; Length 533;
Best Local Similarity 22.5%; Pred. No. 1.3;
Matches 85; Conservative 54; Mismatches 116; Indels 122; Gaps 18;
```



```

QY      14 POOMGHSSFLHRLFGS---LRAMRASSQLLWSEAL----GGFLLAVVYGSAPFVPPS    64
       | : | : : : | : | : : : | : | : : : | : | : : : | : | : : : |
Db     102 PKSGGDYSYVKDIFGGLAGFLRLMTAVLVITYPTNQAVIALTFPSNVLQPLEPFC-FPPES   160
QY      65 ALGIGLAATAAYWALLSTFDID--LRQATPIHWL-----VLLWGVDALATG--   109
       | | | | | | | | | | : | : | : | : | : | : | : | : | : |
Db     161 GLRR-LAAI---CLLLTWVNCSYRWATRVDIFTAGKLALALLIINGVYICKGEF   215
QY     110 --LSPVRA-----AALVGIA-----K                               123
       | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     216 FWLEPKNAFFENFOEPDIGVALAFLOGSFAYGGWNFLNYTEELVDPYKNLPRAIFISIP   275
QY     124 LTLVLLVEFALARV-LRNPRLSILFSVYVITSLSFVSVYGLNWITGYVE-ELATWVDRNS   181
       | : | : | : | | | : | : | : | : | : | : | : | : | : | : |
Db     276 LVTFEYVEFANIAVTAMSP-QELLASNAVAVFEGEKLLGVMAWIMIPISVALSTFGGVNG   333
QY     182 VADFTSRVYSYLGNPNLLAAYLVPTTAFSAAGVWRGWLPKLL-----AIIA    229
       : | : : : | : | : | : | : | : | : | : | : | : | : | : |
Db     334 SLFTSSRLF-----FAGAREGHLPSVLAMIHYKRCTPIPALLE   371
QY     230 TGASSLCLILT---YSRGWGIVAMIF---VMALLGLYWFOPLRPAPWR-RWLFPVVL   281
       | : | : : : | : | : : : | : | : | : | : | : | : | : | : |
Db     372 TCLSTLLMLVTSDMYTLLINYGVFINLYFGVTVAGQIVLRWKKKPDIPRIKISLFFPIY   431
QY     282 GGLVAVLLVAVLGLEPL 298
       | | | : | | | :
Db     432 LRFWAFLLIIFSLWSEPv 448

```

Search completed: January 10, 2002, 23:09:02
Job time: 304 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2002, 23:04:08 ; Search time 56.67 seconds
(without alignments)
1205.385 Million cell updates/sec

Title: US-09-887-038-3
Perfect score: 2426
Sequence: 1 MTVMQTLTFAHYQPQMGHS.....WQPPSKLPPEAEHSDEKM 467

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_17:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2426	100.0	467	2 P72547	P72547 synechococc
2	1132	46.7	474	2 P73958	P73958 synechocyst
3	206	8.5	438	2 Q9KTH1	Q9KTH1 vibrio chol
4	201.5	8.3	387	2 P72672	P72672 synechocyst
5	195	8.0	443	2 Q9PH43	Q9PH43 xylella las
6	180	7.4	478	2 Q9IIM9	Q9IIM9 pseudomonas
7	178	7.3	641	2 Q9ABR1	Q9ABR1 caulobacter
8	172.5	7.1	438	2 Q9KYG4	Q9KYG4 streptomyc
9	170	7.0	391	2 Q9KKE4	Q9KKE4 pseudomonas
10	169.5	7.0	384	2 Q9RZA8	Q9RZA8 deinococcus
11	168	6.9	422	2 Q9A5L8	Q9A5L8 caulobacter
12	166.5	6.9	597	2 Q9KUM6	Q9KUM6 vibrio chol
13	161.5	6.7	440	2 Q9AAU5	Q9AAU5 caulobacter
14	159	6.6	861	2 O06944	O06944 synechocyst
15	154	6.3	360	2 Q9XC97	Q9XC97 klebsiella
16	151.5	6.2	401	2 Q9HUG6	Q9HUG6 pseudomonas
17	151.5	6.2	461	2 Q9I6E8	Q9I6E8 pseudomonas
18	149	6.1	395	2 Q9Z404	Q9Z404 pseudomonas
19	149	6.1	539	2 P74814	P74814 sphingomona

20	144.5	6.0	381	2 Q9ZG09	Q9ZG09 thiobacilli
21	139.5	5.8	388	2 Q9RT25	Q9RT25 deinococcus
22	139.5	5.8	407	2 Q9A5Z6	Q9A5Z6 caulobacter
23	138.5	5.7	402	2 Q9ZIT8	Q9ZIT8 escherichia
24	138	5.7	515	2 Q9RRK1	Q9RRK1 deinococcus
25	137.5	5.7	443	2 Q9KEH4	Q9KEH4 bacillus ha
26	137.5	5.7	474	2 Q9RU99	Q9RU99 deinococcus
27	134	5.5	335	2 Q9KSL2	Q9KSL2 vibrio chol
28	134	5.5	402	2 Q9I2B6	Q9I2B6 pseudomonas
29	133.5	5.5	535	2 Q9RW6	Q9RW6 pseudomonas
30	133.5	5.5	585	2 P72627	P72627 synechocyst
31	131.5	5.4	515	2 Q9KYW0	Q9KYW0 streptomyc
32	131	5.4	524	10 O80905	O80905 arabidopsis
33	130.5	5.4	577	2 Q9L134	Q9L134 streptomyc
34	130	5.4	536	2 Q9L6A1	Q9L6A1 pasteurilla
35	130	5.4	649	2 Q9KE20	Q9KE20 bacillus ha
36	129.5	5.3	405	2 O68200	O68200 escherichia
37	129	5.3	487	1 Q9YG70	Q9YG70 aeropyrum p
38	128.5	5.3	530	2 Q9I726	Q9I726 pseudomonas
39	128.5	5.3	566	1 Q9HS12	Q9HS12 halobacteri
40	128	5.3	461	2 O51524	O51524 pseudomonas
41	127.5	5.3	437	2 Q9I0U5	Q9I0U5 pseudomonas
42	127.5	5.3	678	2 Q9A4E2	Q9A4E2 caulobacter
43	127	5.2	591	2 O53150	O53150 mycobacteri
44	127	5.2	709	2 Q9RYK7	Q9RYK7 deinococcus
45	126.5	5.2	467	2 O31199	O31199 pseudomonas

ALIGNMENTS

RESULT 1
P72547
ID P72547 PRELIMINARY: PRT; 467 AA.
AC P72547;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, last sequence update)
DT 01-AUG-1998 (TREMBLrel. 07, last annotation update)
DE PUTATIVE MEMBRANE PROTEIN.
GN DC14.
OS Synechococcus sp. (strain PCC 7942) (Anacystis nidulans R2).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=1140;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PCC 7942;
RA Bonfil D.J., Lieman-Hurwitz J., Ronen-Tarazi M., Kaplan A.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U62616; AAB08477.1; -
SQ SEQUENCE 467 AA; 51570 MW; 2C3FEB404ED462BF CRC64;

Query Match	100.0%; Score 2426; DB 2; Length 467;
Best Local Similarity	100.0%; Pred. No. 2.2e-123;
Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY 1 MTVMQTLTFAHYQPQMGHSFHLRFLGSLRAWRASQLLVWSEALGFLLAVVYGSAPF 60	
Db 1 MTVMQTLTFAHYQPQMGHSFHLRFLGSLRAWRASQLLVWSEALGFLLAVVYGSAPF 60	
QY 61 VPSSALGLGLAAIAAVYVALLSLTDIDLRQATPIHMLVLLYWGVDALATGSPVRAALVG 120	
Db 61 VPSSALGLGLAAIAAVYVALLSLTDIDLRQATPIHMLVLLYWGVDALATGSPVRAALVG 120	
QY 121 LAKLTLYLVFALFALFARVLRNPRLSLFSVVVITSLFVSVYGLNOMIYVEELATWDRN 180	
Db 121 LAKLTLYLVFALFALFARVLRNPRLSLFSVVVITSLFVSVYGLNOMIYVEELATWDRN 180	
QY 181 SVADFTSRVSYLGNPNLLAAYLVPTTAFSAAGVWRGWLPKLLAATGASSCLILT 240	
Db 181 SVADFTSRVSYLGNPNLLAAYLVPTTAFSAAGVWRGWLPKLLAATGASSCLILT 240	
QY 241 YSRGGWLGFMAMIFVWALLGLYWFQRLPAPWRRLFPVVLGLVAVLLVAVLGLPLRV 300	

```
Db 241 YSRGWMGFVAMIFVWALLGLYWFQRLPAPWRWLFPPVVLGLVAVLLVAVLGLLEPLRV 300
QY 301 RVLSIFVGREDDSSNNFRINWVLAVLQMIQDRPWLIGPGNTAFNLVYPLYQOARFTALSA 360
Db 301 RVLSIFVGREDDSSNNFRINWVLAVLQMIQDRPWLIGPGNTAFNLVYPLYQOARFTALSA 360
QY 361 YSVPLEVAVEGGLGLTAFAMLLVTVAVTRQVSRRLRDRNPQAFWLMASLAGMIG 420
Db 361 YSVPLEVAVEGGLGLTAFAMLLVTVAVTRQVSRRLRDRNPQAFWLMASLAGMIG 420
QY 421 HGLFDTVLYRPEASTLWMLCIGATASFQOPQSKOLPPEAHSDEKM 467
Db 421 HGLFDTVLYRPEASTLWMLCIGATASFQOPQSKOLPPEAHSDEKM 467

RESULT 2
P73958 PRELIMINARY; PRT; 474 AA.
ID P73958
AC P73958;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE HYPOTHETICAL 52.2 KDA PROTEIN.
GN SLR1515.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_Taxid=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; Pubmed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirose M., Sugita M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT *Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
DR EMBL; D90911; BAA18025.1; -.
DR InterPro; IPR001457; Oxidored_q3.
DR Pfam; PF00499; oxidored_q3; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 474 AA; 52241 MW; 017A49CBC2758EC6 CRC64;

Query Match 46.7%; Score 1132; DB 2; Length 474;
Best Local Similarity 47.5%; Pred. No. 8.2e-54;
Matches 223; Conservative 79; Mismatches 163; Indels 4; Gaps 2;

QY 1 MTVMQTLTFAHYQPOQWGHSSFLHRLFGSLRAWRASSQLVWSEALGGFLAVVYGSAPF 60
Db 5 ISIMRSLMEGGFSPQEWGRGSLVHRLVWGQSWIQASVLWPHFEALGTALVAIIIFIAPF 64
QY 61 VPSSALGLGLAIAIAYWALLSTLDIDLRQATPIHMLVLLYWGVDALATGLSPVRAALVG 120
Db 65 TSTMTGLIFMLLCGAFWALLTFADQPGKGLTPIHVLVFAWCISAIAVGFSPVKMAASG 124
QY 121 LAKTLFLLVLFALARVLRNPRLSLFSVVVITSLFVSYYGLNQWITYGVEELATWDRN 180
Db 125 LAKTLRNLCLFLLARLLQNKQWLNRLVTVVLLVGLVGSYGLRQQVDVEQLATWMDPT 184
QY 181 SVADFTSRVSYLGNPNLLAAYLVPTAFSAAAGVWRGWLKLLAIAATGASSCLILT 240
Db 185 STLAQATRVYSFLGNPNLLAAYLVPTGLSLSALVWWRWMPKLLGATMVIYNLLCLEFT 244
QY 241 YSRGWMGFVAMIFVWALLGLYWFQRLPAPWRWLFPPVVLGLVAVLL--VAVILGLEPL 298
Db 245 QSRGWMGLAVLALGATFLALCYFWMLPQLPKFWQWMSLPLATA--VAVILGGALLAVEPT 302
QY 299 RVRVLSIFVGREDDSSNNFRINWVLAVLQMIQDRPWLIGPGNTAFNLVYPLYQOARFTAL 358
Db 299 RVRVLSIFVGREDDSSNNFRINWVLAVLQMIQDRPWLIGPGNTAFNLVYPLYQOARFTAL 358
```

```
Db 303 RLRAWSIFAGREDDSSNNFRINWEGVKAMIRAPIIGIGPGNEAFNOITYPMRPTAL 362
QY 359 SAYSVPLEVAVEGGLGLTAFAMLLVTVAVTRQVSRRLRDRNPQAFWLMASLAGMIG 418
Db 363 SAYSVILEIVETGVGFTCMWLLAVTLGKVELVKRCRQTLAPEGTIWIGALAAIIGL 422
QY 419 LGHGLFDTVLYRPEASTLWMLCIGATASFQOPQSKOLPPEAHSDEKM 467
Db 423 LVHGMVDTWYRPVSTLWMLLVAIVASQWASQAARLEASKENEDEKPL 471

RESULT 3
O9KTH1 PRELIMINARY; PRT; 438 AA.
ID O9KTH1
AC O9KTH1;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, last annotation update)
DE HYPOTHETICAL PROTEIN VC0931.
GN VC0931.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_Taxid=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; Pubmed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gili S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermoлаeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RL Nature 406:477-483(2000).
DR EMBL; AE004176; AAF94093.1; -.
DR TIGR; VC0931; -.
KW Complete proteome.
SQ SEQUENCE 438 AA; 49177 MW; FDF34AAC184C429B CRC64;

Query Match 8.5%; Score 206; DB 2; Length 438;
Best Local Similarity 24.7%; Pred. NO. 0.00046;
Matches 120; Conservative 68; Mismatches 200; Indels 98; Gaps 19;

QY 5 QTLTFAHYQPOQWGHSSFLHRLF-----GSLRAWRASSQLVWMS--EALGGFLAVVYG 56
Db 6 KTLFPA-----FLFVFWLPIPLGSNRPW-----AWSLNELTVAFL----- 42
QY 57 SAPFVPSSALG-LGLAIAIAYWALLSTLDIDLRQATPIHMLVLLYWGVDALATGLSPVRA 115
Db 43 -----LGCFLFLLPAAQCWLAIQRAKAVIL---IPVALFTLMSWMSQWPSLGLSSDQT 90
QY 116 AALVGLAKTLTYLLVFALARVLRNPRLSLFSVVVITSLFVSYG----- 162
Db 91 LVFISAVKSLHYLQICLVALLVDTPHKLMKMTATTMIASGICQAFYAGVILLLELQTSPT 150
QY 163 ----LNQWITYGVEELATWDRNSVADFTSRVSYLGNPNLLAAYLVPTTAFSAAAGVWR 218
Db 151 FSLPLNORASG----SPVYHNHLANFL-MNMLCLGFGLLIAELNHQTTQ-----GWR 197
QY 219 GWLPKLLAIAATGAS-----SLCLLYTSRGGLGVAMIFVWALLGLYWFQRLP 269
Db 198 NTLKLLAAILLSDKAFIRGLVIMVIALVTRSRMGNTAFVALSIGSLILFLFYKSK-- 255
QY 270 APWRWMLFPVVLGLVAVLLVAVLGLLEPLRVLSIFVGREDDSSNNFRINWVLAVLQMIQ 329
Db 256 -PKSLYLLIVSLFVADALVSNMFGLDKVRQR-----LAETSLQNESRDVDVADALNAIQ 309
QY 330 DRPWLIGPGNTAFNLVYPLYQOARFTAL--SAYSVPLEVAVEGGLGLTAFAMLLVTV 387
Db 330 DRPWLIGPGNTAFNLVYPLYQOARFTAL--SAYSVPLEVAVEGGLGLTAFAMLLVTV 387
```

[illegible][illegible]

RESULT	5		
Q9PH43	ID	PRELIMINARY;	PRT; 443 AA.
Q9PH43	ID	Q9PH43;	
AC	01-OCT-2000 (TREMBLrel. 15, Created)		
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)		
DT	01-MAR-2001 (TREMBLrel. 16, Last annotation update)		
DE	MEMBRANE PROTEIN.		
GN	XF0103.		
OS	Xylella fastidiosa.		
OC	Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;		
OC	Xylella.		
OX	NCBI_TaxID=2371;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=9A5C;		
RX	MEDLINE=20365717; PubMed=10910347;		
RA	Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,		
RA	Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,		
RA	Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,		
RA	Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carer H.,		
RA	Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,		
RA	Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,		
RA	Facchini A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,		
RA	Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,		
RA	Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,		
RA	Ho P.L., Hohelsel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,		
RA	Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,		
RA	Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,		
RA	Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,		
RA	Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,		
RA	Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,		
RA	Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,		
RA	Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,		
RA	de Oliveira M.C., de Oliveira R.C., Palmeri D.A., Paris A.,		
RA	Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,		
RA	Queaglio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,		
RA	de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,		
RA	da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,		
RA	da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,		
RA	de Souza A.P., Terenzi M.A., Verjovski-Almeida S., Vettore A.L.,		
RA	Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,		
RA	Zago M.A., Zatz M., Meldanis J., Setubal J.C.;		
RT	"The genome sequence of the plant pathogen Xylella fastidiosa.";		
RL	Nature 406:151-159(2000).		
KW	EMBL; AE003864; DAF82916.1; -.		
KW	Complete proteome.		
SEQUENCE	443 AA; 48239 MW; E53A1CE9A66FB56F CRC64;		

Query Match	8.0%;	Score 195;	DB 2;	Length 443;
Best Local Similarity	25.2%;	Pred. No. 0.0018;		
Matches 107;	Conservative 47;	Mismatches 158;	Indels 112;	Gaps 21;

QY	56	GSAPFVPSALGLGLAIAAYW-----ALLSLTDIDLRQATPIHWLVLLYWGVDATA	107
		: : : : : : : : :	
Db	63	GGTQLDPSPAWALTSLFMAYWLPQVSLFGALELSESLRKVATDRLYLPFMW--LCAIA	120
		: : : : : : : : : :	
QY	108	TGTSFVRAALVGLAKLTLYLLFEALARVLRNPRLSLFSVVVITSLEFSVYGLNQWI	167
		: : : : : : : : : :	
Db	121	VGSDERERKRTFKGLAVIGLIWTFDALAQAMLHT-----SPLFWSLNQLKQAI	167
		: : : : : : : : : :	
QY	168	YGVEEIATWVDNRNSVADFTSRVSYLGNPNLLAAYLVPTT-----AFSAAATGVWRGWL	221
		: : : : : : : : : :	
Db	168	SGYFCST--RQMI--LADRLSGIFGPCNLKLGOTLATLSFVLFAMQKGLAW----	218
		: : : : : : : : : :	
QY	222	PKLLIAATGASSLCILITYSRGWLGF--VAMIFVWALLGLYWFQRLPAPWRRWLPV	279
		: : : : : : : : : :	
Db	219	--LIVVVLG---IVLLAGSRASWITYALILMLSGWRVIGT-----RRLI---	259
		: : : : : : : : : :	
QY	280	VIGGLVAVLVAVLGLLEPL--RV-RVLSIFVGR--EDSSNNFRINWVLAVLQMIQDRP	332
		: : : : : : : : : :	
Db	260	GMGALFLPALAVIAFSPQTRERIDRTAAVFADHGAGVDQALSGRSQIWOAALCMIQAH	319
		: : : : : : : : : :	

QY 333 WLIGPGNTAFNLVY-----LYQARFTALSAYSVPLEVAVEGGLGL-----TA 378
Db 320 LSGVGV--RCFBRAYPACNPTPERIPAWGAGPALHAQIVLEILSETCVIGLLMIAGAA 377
QY 379 FAWLLLVTAFTAVRQVSRLRDRNPQA-----FW--LMASLAGL- 415
Db 378 MAW-----RAMRYATAAARDRAPAMISLLATVEFPNTHLAFYSSFGALMLMLAGLY 430
QY 416 AGML 419
Db 431 AGAL 434

RESULT 6
Q91IM9
ID Q91IM9 PRELIMINARY; PRT: 478 AA.
AC Q91IM9;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE HYPOTHETICAL PROTEIN PA2240.
GN PA2240.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen."
RL Nature 406:959-964(2000).
DR EMBL; AE004649; AAG05628.1; -
DR InterPro; IPR003439; ABC_transportr.
DR PROSITE; PS00211; ABC_TRANSPORTER; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 478 AA; 52495 MW; 6C4F36DABAEF70F1 CRC64;

Query Match 7.4%; Score 180; DB 2; Length 478;
Best Local Similarity 23.0%; Pred. No. 0.013;
Matches 124; Conservative 67; Mismatches 159; Indels 188; Gaps 29;
QY 48 GFLAVVYGSA-----PFPSSALGLGLAIAA-----YWALL-----SL 82
Db 4 GNLAVVYGLAFGGALLMLSPAKAGAA-VGLAAVTLRPPFWGILLPALVATFMPYST 62
QY 83 TDIDLRQATPIHMLVLLYGV--DALATGLSPVRAAALVGLAKLTLYLVFA----- 132
Db 63 VNVGIR-STVSEAILALTWGAVLWHIFLSRLPPAPALRYRSTERMLLMLFTVLPVVG 121
QY 133 -----LAARVL-----RNPRLSLLFSVVITSLFVSIV- 161
Db 122 QVSIKAEASGLSNWLRLNLSIVELAGRLVERKNRETLVIALLLGLTAMLSIAVFI 181
QY 162 -----GLNQWLYGVEELATWVDRNSVADFTSRVSYLGNPNL---LA 200
Db 182 RYRSASGMAPILAMFNANLDTLKFGLLEA-----SSRMGSPWHPNATGIM 229
QY 201 AYLVPPTTAESAAGVWRGWLPRKLLAIAATGASSLCLILYSRGWLGVAMIFWALLG 260
Db 230 ALLPLAFCYGANOGWRRGLG--LAVAVLGAAA--ILLASSRGAMLSLAVVLF-WMSLR 284
QY 261 LYWFQRLPAPWRRLFPVVLG-GLVAVLLVAVLGLPLEPRVLSIFVGREDSNNPRIN 319

Db 285 KVPYTGRL-----LLLGVLVVL---VLSYPIQLERLATIF-SPONASTEVRFD 330
QY 320 VMLAVLQMIQDRPWLIG-----PG-----NTAFNLVYPLYQARFTALSAYSVPL 365
Db 331 EYRMFPKAVARYP-IGIGFKYDPPVPGTDLIGSINIMINFWY----- 371
QY 366 EVAVEGGLIGLTAFAWLLLVTAFTAVRQVSRLRDRNP-----QAFWLMASLAGL 419
Db 372 KVGIL-GGMLEFIATW-----RWMREARPEKPIRLTRDNAIWL-GSTGILAL 419
QY 420 GHGLFDTVL-YRPEASTLWMLCIG-----AIASFWQPP-----SKQLPEAE 461
Db 420 VSGLFDPHYEFAVVMIGLEFNLVGINLLLEARRLFFPERQPPRAVGRKLRQLERGAE 477

RESULT 7
Q9ABR1
ID Q9ABR1 PRELIMINARY; PRT: 641 AA.
AC Q9ABR1;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE HYPOTHETICAL PROTEIN CC0165.
GN CC0165.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=69394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Utterback T., Tran K., Wolf A., Yamalhevyan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE005690; AAK22152.1; -
DR TIGR; CC0165; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 641 AA; 67175 MW; D8FF63BE76B565F9 CRC64;

Query Match 7.3%; Score 178; DB 2; Length 641;
Best Local Similarity 23.1%; Pred. No. 0.021;
Matches 115; Conservative 80; Mismatches 193; Indels 110; Gaps 23;
QY 18 GHSSFLHRLFGSLRAWRASQLLVWSE--ALGGFLAVV-----YGSAPRPSSALGLGL 70
Db 10 GSSS-----FGS-----SGSRLKVSSEGAIGALLTLILAETAFGASELAVSSA----F 55
QY 71 AAIAYWALLSITDID-----LRQ-ATPIHML-----VLLYGVDAATG-----L 110
Db 56 AALHALFPVLLATCGMARNAARQGVSPITMLPFAIIMLVTAWALTPFGPGPHPVLYV 115
QY 111 SPVRAAALVGLAKLTLYLL-----VFALAARVL--RNPRLSLLFSVVITSLFVSIVG 162
Db 116 EPNCGAITVDRSSLIINLVRLGGLACFLAQAQIGSSSETRRALFWLSLVALGAY-AIIA 174
QY 163 LNQWLYGVEELATWVDRNSVADFTSRVSYLGNPNLLAAYLVPTTAESAAGVW----- 217
Db 175 ILQHV-GVRA-----SERMTATLSPNTLATLMGVGAVFAAMEVTOFIORAG 220
QY 218 -RGWLPKL-----LAIATGASSLCLILYSRGWLGVFV--AMIFWALLGLWFOPL 268
Db 221 GRLSLDKPLDAGISLALLAVFAVALAFTGSGVEFATVVALAIFLWVYL----- 271
QY 269 PAFWRRLFPVVLGGLVAVLLVAVLGLPLEPRVLSIFVGREDSNNPRINWMLAVLQMI 328
Db 272 -AQGRARVVMIGG--AAVLLAAVGLAMRSADITTAARLENLDGDVATRTQITFAAHMQAF 328

OY 329 QDRPWLIGIPGNTAFNLVPLYQO-----ARFTALSAYSVPLEVAVEGGLGLTAF 379
: ||| | | : | : | : | : | : | : | : | :
Db 329 KASPMFGYGLGS-----FPAYNQWMTSETLPALFDVRAHNLVLYLWLEAGLVGALISM 382
OY 380 AMLLVTAVTAVROVSRLRRDRNPQAFWLMASTAGLGHGLEDVLYRPEASTLWML 439
| | | | | : | : | | | : | : | : | : | :
Db 383 LGILGV---LIMOVCAANKPCTSGALARAIASTALVLDHMSDFAVQAPALQALFAV 438
OY 440 CIGATASFQWQPSKOLP 457
: || : : | : | :
Db 439 GLGALVTS-SPARGRLP 455

RESULT 8
O9KYG4 PRELIMINARY; PRT; 438 AA.
ID O9KYG4
AC O9KYG4;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE PUTATIVE INTEGRAL MEMBRANE PROTEIN.
GN SCC61A.29.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycetales; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Oliver K., Harris D.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP 'SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP 'SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kleser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL356595; CAB92274.1; -.
SQ SEQUENCE 438 AA; 45017 MW; 38173961751FEFA1 CRC64;

Query Match 7.1%; Score 172.5; DB 2; Length 438;
Best Local Similarity 26.6%; Pred. No. 0.029;
Matches 119; Conservative 63; Mismatches 178; Indels 87; Gaps 23;

OY 50 LLAVVYGS-----PEVPSSALGL---GLAATAAYWALLSLTDIDLRQATPIHMLVLYW 101
: ||| | | : | : | : | : | : | : | : | :
Db 19 VLAVVAVVALLALPLAPGGGGAHPADAVSALAVLYCALRLVR-PLSRTAAVVL 75
OY 102 GVDALATGLSPVRAAL--VGLAKLTLYLVFALA-ARVLRNPR-----LRSLFSVVI 153
: | | | : | : | : | | | | | | : | : | :
Db 76 GLPVL--GLAATAAGAVSPGAGLAGRITQYFVLVPAVLLVRDRADYRLLAWSMAGL 133
OY 154 TSLFVSVYGLNQWITYVEELATWDRNSVADFTSKVYSYLGPNLLAAYLVPTTAFSAA 213
: | : | : | : | : | : | : | : | : | : | :
Db 134 -ALMOGAVGVHQYVTGTG--ASYQGERIRAVGTGPDVDMGMATVVSGLV-----CA 183
OY 214 IGVWRGWLPKLLAIAATGASSICLI--LTYSRGWLGFWAMIFWALIGLWFOPLPA 270
: | : | : | : | : | : | : | : | : | : | :
Db 184 VGLALGRTPVQRILFAAGCALVLLPLALSPSRGAWIATAVTCAYQDLL-----A 233
OY 271 PWRRWLFPPVVLGIVAVLLVAVLGL--EPLRVVLSI--FVGREDDSSNNFRINWLVAVLQ 326
: | | | : | : | : | : | : | : | : | : | :
Db 234 GVRRAI AVGAANA AAVLVVVGFGVGTAMIQERVGSITQVADAPDQSVTDRTYTMMAAAG 293

OY 327 MIQDRPWLIGIPGNTAFNLVPLYQO-----RFTALSAYSVPLE 366
: ||| | | : | : | : | : | : | : | : | :
Db 294 MWRERPLTVGLKG-----FPEHRDAHSLALSSGSETDGAGAGYRRQPLSPHMYLL 347
OY 367 VAVEGGLGLTFAF--WLLVTAVTAVROVSRLRRDRNPQAFWLMASTAGLIA-GMLGHGL 423
| | | | | : | : | | | : | : | : | : | :
Db 348 VIAEQGLIGLALAGSMIALI--VLGLRLRAARR-----AHGAVPDCAFVACGLLVWL 400
OY 424 FDTVL--YRPEASTLWLCIGATASF 448
| : | : | : | : | : | : | :
Db 401 TDEAVADIGPSTVLTAFCG-LAAMW 426

RESULT 9
O9KKE4 PRELIMINARY; PRT; 391 AA.
ID O9KKE4
AC O9KKE4;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE O-ANTIGEN POLYMERASE.
GN WZY.
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KT2442;
RX MEDLINE=21128631; PubMed=11207769;
RA Rodriguez-Herva J.J., Reniero D., Galli E., Ramos J.L.;
RT "Cell envelope mutants of Pseudomonas putida: physiological
RT characterization and analysis of their ability to survive in soil.";
RL Environ. Microbiol. 1:479-488(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=KT2442;
RA Rodriguez-Herva J.J., Reniero D., Ramos J.L.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF109736; AAF15969.2; -.
SQ SEQUENCE 391 AA; 43109 MW; 4658DF8489DB64B0 CRC64;

Query Match 7.0%; Score 170; DB 2; Length 391;
Best Local Similarity 23.1%; Pred. No. 0.036;
Matches 102; Conservative 63; Mismatches 157; Indels 120; Gaps 23;

OY 42 WSEALGFLLAVVYGSAPFVPS-----ALGLAATAAYWALLSLTDIDLRQATPIH 94
: ||| | | : | : | : | : | : | : | : | :
Db 7 WAQAMTIGLVWFELAIALAPSNKYQGLVFLWPLTVLAWSARGVL-VQAWRRQPAL 65
OY 95 W---LVLLYGVDAATGLS-----PVRAALVGLAKLTLYLVFALAAVLR---N 140
| | | | | : | : | : | : | : | : | : | :
Db 66 WGSVLLDLAW-----SGLSLAWSDAEPMRE-----LKRLLYILVFLAFLAQLQIQ 113
OY 141 PLRSLLF--SVVITSLFVSVYGLNQWITYVEELATWDRNSVADFTSRVYS--YLGNP 196
: | | | | : | : | : | : | : | : | : | :
Db 114 ARIRQLLELGSALLAIALVSIIR---FYGVQRAPLFLRAGIGEISHPILAGVYIGSA 169
OY 197 NLLAAYLVPTTAFSAAAGVWRGWLPKLLAIAATGASSICLITLTSRGWLGFWAMIFW 256
: | | | : | : | : | : | : | : | : | : | :
Db 170 LILMLYEPPRRR-----GLQLW---LAALACIGAFAM---LSQSGAVLALVITVMA 217
OY 257 ALLGLWFPRLPAPWRRWLFPPVVLGGLVAVLLVAVLGLLEPLRVKLSIFVGREDDSSNNF 316
: | : | : | : | : | : | : | : | : | : | :
Db 218 PL-----WFRDRHSR-----VFSVALLATGLAFLAVYDVIAQRSSSY 255
OY 317 RINWVLAVLQMIQDRPWLIGIPGN-----TAFNLVPLYQOARFTALSAYSVPLEVAVEG 372
: | : | : | : | : | : | : | : | : | : | :
Db 256 RPEIFHAYVQMIAAHPWTGLGLGADYEVSAGVMHFD-HTHNMET-----HVAVEMG 305
OY 373 LLG--LTAFAWLLLVTAVTAVROVSRLRRDRNPQAFWLMASTAGLGLGHGLFDVLYR 430

Db 306 LPMGLLVWVWVLTGLGEIVRAR-----GTLFGKV-LLGFWYSTLAMQ 347
QY 431 PEASTL-----W--WLCIG 442
Db 348 FDAASLTGTPRAEWFTSWLPVG 369

RESULT 10
Q9RZA8 PRELIMINARY; PRT; 384 AA.
AC Q9RZA8;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE HYPOTHETICAL 41.9 KDA PROTEIN.
GN DRA0046.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1;
RX MEDLINE=20036896; Pubmed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
radiodurans R1.";
RL Science 286:1571-1577(1999).
DR EMBL; AE001862; AAF12372.1; -
DR TIGR; DRA0046; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 384 AA; 41905 MW; 744B3C6B599EF128 CRC64;

Query Match 7.0%; Score 169.5; DB 2; Length 384;
Best local Similarity 26.7%; Pred. No. 0.037;
Matches 113; Conservative 56; Mismatches 158; Indels 97; Gaps 24;

QY 50 LLAIVYGSAPFVPSALGLGLAATAVWALLSLTDIDLRQATPIHMLVLLWGVDAIATG 109
Db 21 LIALV-----PVFP-----LYLAALACLSLRLP-----QAAR---GVLEFFRAATQILIAA 64
QY 110 LSPVRAALVGLAKL-TLYLLVFAALAVLRNPR-LRSLFSVVVITSLFVSAYGLNQMI 167
Db 65 LFTPAPLISVGLAARTLLILANVAAGVYLRDSRLRLPLMLGQLI---FVTA-----WI 116
QY 168 YGVEELATWDRNSVADFTSRV---SYLGNPNLLAAYLVPTTAFSAAGVWKGWLPKL 224
Db 117 Y---TIST-----QGVAGVQERLGHPIYYIVSLGLVAVVALMIIVFWRGAAMWR-WPAGL 168
QY 225 LAIAATGASSLCLILTYSRGWLGFVAMIFWVALLGLYWFQRLPAPWRWR-LFPVVLGG 283
Db 169 LALA-----TFAAGSRGLLALGVGLAALAFGGQRR---RWVWMLPAVLVM 213
QY 284 LVAVLLVAV-LGLEPLRVRLSIFVGREDDSSNNFRINWVLAQLMIQDRPWIGIGPNTA 342
Db 214 LAFAATTSINVPFKPLRLINDQTSRE-----YVWQDAVSGWETSPLGSGVPYQGG 265
QY 343 FNLVY-----PLYQO-----ARFTA--LSAVSVPLEVAVEGGLGLTAFAWL 382
Db 266 PYLTLYLFKDCQLTPTLQRNKIECPQLSRWSSVWLIAHNAMLHMLLESGLIGLSG----- 321
QY 383 LVAATAVAVRQVSRRLRDRNPQAFWLMASLAGMLGHGLFDVLY--RPEASTLWMLC 440
Db 322 LLAIMVYALMRAIQL---GDP-----FTLAVLYGFTAMVVVDVIAVPSHPHSEIMWVC 372
QY 441 IGAI 444

Db 373 VGLV 376

RESULT 11
Q9A5L8 PRELIMINARY; PRT; 422 AA.
AC Q9A5L8;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE HYPOTHETICAL PROTEIN CC2429.
GN CC2429.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=69394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21173698; Pubmed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
Potocka I., Nelson W.C., Newton A., Stephens C., Pladke N.D., Ely B.,
DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE005912; AAK24400.1; -
DR TIGR; CC2429; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 422 AA; 45786 MW; 656279D26E34E1FF CRC64;

Query Match 6.9%; Score 168; DB 2; Length 422;
Best local Similarity 24.9%; Pred. No. 0.049;
Matches 94; Conservative 57; Mismatches 150; Indels 76; Gaps 17;

QY 44 EALGPELLAVVYGSAPFVPSALGLGLAATAVWALLSLTDIDLRQATPIHMLVLLWGV 103
Db 50 EAASGLL-----RVAFLEAYGLAF-LLIMGRPMNVVVT--VRQPLILMVV--V 96
QY 104 DALATGLSP--VRAALVGLAKLTYLLVFAALAVLRNPRLSLFSVVVITSLFVS 160
Db 97 ASISWSIQPDVSIKGFANCTP-TLAGLALASHFRWAEARLFAIYYCFILIVASYLSL 154
QY 161 ---YGLNQVIYGEELATWDRNSVADFTSRVSYLGNPNLLAAYLVPT-----TAFSA 211
Db 155 ALPSIGVMTIEIFGAWRGIMMEKNGLGIMAFGACLLG---AAALNPDRAKLFTLFA 210
QY 212 AAGVWRGWLPKLLAIAATGASSLCLILTYSRGWLGFVAMIFWVALLGLYWFQ 266
Db 211 LAIGL-----VLLSQSKVALASLL-----LGMTALGFVWIVQRPALG----- 248
QY 267 RLPAWRRLFPVVLGLVAVLLVAVLGLPELRVRLSIFVGREDDSSNNFRINWVLA 326
Db 249 -AAATW-----AGVYGALLLAMFVL--LASDMLFEILGR-DATLTGRTEIWTAA 295
QY 327 MIODRPWLGIG-----PGNTAFNLVVPPLYQOARFTALSAVSVPLEVAVEGGLGLTA 381
Db 296 QIEQRPWQGYGYAAVWSKSGWGPLWIVNDAKFVPOHAHNSWLEQNLGIGLGLIANG 355
QY 382 LLVTAATAVAVRQVSRRLR 398
Db 356 FYLQTMSLAVVAVYRER 372
RESULT 12
Q9KUW6 PRELIMINARY; PRT; 597 AA.
ID Q9KUW6
AC Q9KUW6;
DT 01-OCT-2000 (TREMBlrel. 15, Created)

DT 01-OCT-2000 (Tremblrel. 15, last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, last annotation update)
DE HYPOTHETICAL PROTEIN VC0393.
GN VC0393.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
cholerae.";
RL Nature 406:477-483(2000).
DR EMBL; AF004127; AAF93566.1; -
DR TIGR; VC0393; -
DR InterPro; IPR000215; Serpin.
DR PROSITE; PS00284; SERPIN; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 597 AA; 67451 MW; D33513A00836688C CRC64;

Query Match 6.9%; Score 166.5; DB 2; Length 597;
Best local similarity 24.3%; Pred. No. 0.083;
Matches 109; Conservative 60; Mismatches 185; Indels 95; Gaps 22;

QY 49 FLAVVYGSAP--FVPS-SALGLGLAAIAVWALLSLT-----DIDLQATPIHWLVLL 99
DB 29 FALAVVYLLAMHFEMPNPGAGLALSFMTWTATISITLAIGLYQLANNQALRYSKLTIG 88
QY 100 YWGVDAATGSLSPV-----RAALVGL-AKLTYLLVFALAAVLRNPRLSLFS 149
DB 89 LMLSCALIT--LPITYSNADWFHTLGRVGLMAGFTLFVLQOF--RFSNKHKQRLMF- 143
QY 150 VVVTISLFSVYGLNQMIYVEELATWDRNSVADF--TSRVYSTLGNPNLLAAYLVPT 206
DB 144 -ILGSLVIAVIGLIQYF-----WLEPGNPGYDTEANRPYGIQGPVNMASFATG 194
QY 207 TAFSAAATGVRGWL-----PKLLAIAATG-ASSLCILLTYSRGWIGFVAMI 253
DB 195 LVLS-----GYLLARQPEKYSRSKIGILYATPLTAPLAVLASRTGWLASLSI 246
QY 254 FVWALLGLYWFQRLPAPWRRLFPVVLGGIVAVLLVAVLGLLEPLRVLSIFVGREDS 313
DB 247 ---ALVAPYLYRFASRRRFTLMLATLG--ICAGFTAMYG-----QGTSGFVANKADL 295
QY 314 NMFRIWMLAVLQMIQDRPWLIGPGNTAFNLVYPLYQARFTALSAYSVPLEV----- 367
DB 296 ESPRRFTFPOTIDMIIEKPTGYGGR--FEAQYLLYTARQHOLNSSYPAGLASMDHPN 353
QY 368 -----AVEGGLGLTAFAMLLVTAVTAVRQVSRLRRDRNPQAFWMLASLAGMLGHG 422
DB 354 ELLYMGVYEGGLPYMG---ILAACFCALRIYAAKRGR-----MAMLLFVPTTLHA 403
QY 423 LFDVLYRPEASTLWMLCIGATSFQPO 451
DB 404 QLEVPFYH---SAIHWTIF-ILITYWVDQ 428

RESULT 13
Q9AAU5 PRELIMINARY; PRT; 440 AA.
AC Q9AAU5;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, last annotation update)

DE HYPOTHETICAL PROTEIN CC0499.
GN CC0499.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OX NCBI_TaxID=69394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwin M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE005722; AAK22486.1; -
DR TIGR; CC0499; -
DR KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 440 AA; 47613 MW; D90CFCDA3B6C5C5F CRC64;

Query Match 6.7%; Score 161.5; DB 2; Length 440;
Best local similarity 26.2%; Pred. No. 0.11;
Matches 95; Conservative 47; Mismatches 143; Indels 77; Gaps 17;

QY 70 LAIAAVWALLSLTDIDLQATPIHWLVLLYWGVDALATGLSP---VRAALVGLAKLT 126
DB 94 LAVLTRSWLLGLLSI-----AVCSVLSLQPDVTIRRLVAVVFTLTG 137
QY 127 YLLVFALAAVLRNPRLSLFS--VVVTISLFSVYGLNQMIYVEELATWDRNSVAD 184
DB 138 VLL-----AERFEMPKTLEVLASTYAVVYTLSE--VFGLFMPAYGVMSI-----D 180
QY 185 FTSRVYSTLGNPNLLA-AYLVPTTAFSAAATG-----VNRGWLPKLLAIAATGASSICL 237
DB 181 FPGAMRGVYGFKNQLGWAMSLATPTFLACALANPQRRRLMGF-----AGAILLI 231
QY 238 ILTYSRGWLGFWAMIFWALLGLYWFQRLPAPWRRLFPVVLGGVAVLV-AVGLF 296
DB 232 VLSTSKTALVSCIAGLAFMPLIALCRIGPALGT-----TALLAASVVALVCAVYALA 284
QY 297 PLRVYLSIFVGREDSNNFRINWMLAVLQMIQDRPWLIGPGN-----TAFNLVYPLYQ 351
DB 285 PELVPEL---TGR-DATFTGRTLMDAISROIQRPLTGYGYGAVWDDLSGWGPVAMISN 340
QY 352 QARFTALSAYSVPLEVAVEGGLGLTAFAMLLVTAVTAVRQVSRLRRDRNPQAFWMLAS 411
DB 341 DQGFRTFSAHNTVLGVWLELGLIGVT--AWMLLVGAM-LKGLSRL--ATAPAYFFLPF 394
QY 412 LA 413
DB 395 LA 396

RESULT 14
O06944 PRELIMINARY; PRT; 861 AA.
AC O06944; Q55995;
DT 01-JUL-1997 (Tremblrel. 04, Created)
DT 01-JUL-1997 (Tremblrel. 04, last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, last annotation update)
DE HYPOTHETICAL 96.7 KDA PROTEIN.
GN SI0737.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96127529; PubMed=8590279;
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 10, 2002, 21:36:02 ; Search time 108.09 Seconds
(without alignments)
11135.948 Million cell updates/sec

Title: US-09-887-038-2
Perfect score: 1404
Sequence: 1 atgactgtctgcaactct.....attcagacgaanaatgtag 1404

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	47.4	3.4	810	22	AAH52001	Mycobacterium tube
2	47.4	3.4	2232	16	AAQ78915	Mycobacterium bovi
3	47.4	3.4	2232	16	AAQ75519	Mycobacterium bovi
4	47.4	3.4	2232	18	AAT99224	M. bovis p55 opero
5	47.4	3.4	2609	22	AAH02108	Mycobacterium tube
6	47.4	3.4	3120	16	AAQ75518	M. tuberculosis in
7	47.4	3.4	3120	16	AAQ78914	M. tuberculosis in
8	47.4	3.4	3120	18	AAT99223	M. tuberculosis in
9	47.4	3.4	3120	19	AAV20433	Mycobacterium tube
10	47.4	3.4	3120	20	AAAX21816	M. tuberculosis in
11	47.4	3.4	3120	20	AAV68212	inhA gene nucleoti

C	12	44.4	3.2	10732	21	AAA10594
	13	39.6	2.8	1723	18	AAT99220
	14	39.6	2.8	1723	18	AAT99219
	15	39.2	2.8	58857	21	AAA58471
C	16	37.8	2.7	6741	21	AAA10595
	17	36.8	2.6	2631	20	AAAX34135
	18	36.8	2.6	2664	20	AAAX34136
C	19	36.2	2.6	1450	12	AAQ13408
C	20	35.8	2.5	1091	22	AAI61038
C	21	35.8	2.5	1124	22	AAI59252
	22	35.8	2.5	2718	21	AAZ60909
C	23	35.6	2.5	10211	19	AAV62152
C	24	35.6	2.5	117213	19	AAV62176
	25	35.2	2.5	390	19	AAV23486
	26	35.2	2.5	390	21	AAA13900
C	27	35.2	2.5	434	21	AAA78189
C	28	35.2	2.5	434	22	AAI28927
	29	35.2	2.5	3978	21	AAC55785
	30	35.2	2.5	4100	20	AAZ32019
	31	35.2	2.5	4100	22	AAC90076
	32	35.2	2.5	12249	21	AAC55840
	33	35.2	2.5	17612	19	AAV23494
	34	35.2	2.5	17612	21	AAA13905
	35	35.2	2.5	17612	22	AAF30870
	36	35.2	2.5	18331	21	AAC55857
	37	35	2.5	30001	18	AAT61016
	38	35	2.5	30001	20	AAAX05110
	39	35	2.5	114955	20	AAAX53491
	40	34.8	2.5	43280	18	AAT80413
C	41	34.6	2.5	1450	12	AAQ13227
	42	34.6	2.5	3058	18	AAT77137
C	43	34.6	2.5	3132	22	AAS06335
	44	34.6	2.5	3646	21	AAZ36714
	45	34.2	2.4	678	22	AAH34854

ALIGNMENTS

RESULT 1	
AAH52001	
ID AAH52001 standard; DNA; 810 BP.	
XX AC AAH52001;	
XX DT 04-SEP-2001 (first entry)	
XX DE Mycobacterium tuberculosis potential drug target gene SEQ ID 55.	
XX KW Drug target; growth; organism viability; characterisation; ds.	
XX OS Mycobacterium tuberculosis.	
XX PN WO200135317-A1.	
XX PD 17-MAY-2001.	
XX PF 13-NOV-2000; 2000WO-US31152.	
XX PR 12-NOV-1999; 99US-0165086.	
XX PR 12-NOV-1999; 99US-0165124.	
XX PR 01-FEB-2000; 2000US-0179531.	
XX PA (REGC) UNIV CALIFORNIA.	
XX PI Eisenberg D, Rotstein SH, Marcotte EM;	
XX DR WPI: 2001-329193/34.	
XX DR P-PSDB; AAG81150.	
PT Identifying nucleotide or polypeptide sequence for use as drug target,	Gene encoding a su
PT Involves providing algorithm that analyzes a functional relationship	M. bovis INH resis
PT between nucleotide or polypeptide sequences, and comparing the	M. tuberculosis IN
	Nucleotide sequenc
	Gene encoding a su
	Mycobacterium spec
	Mycobacterium spec
	Pseudomonas virus
	Human polynucleoti
	Human polynucleoti
	Murine delta-relat
	HSV-2 strain SB5 C
	HSV-2 strain SB5 C
	Pseudomonas Xcpv s
	Pseudomonas alcali
	CDNA encoding huma
	Colon tumour relat
	Type I polyketide
	Human METH1 relate
	l13855 cDNA clone.
	Complete Mitomycin
	Pseudomonas Xpc, O
	Pseudomonas alcali
	Pseudomonas alcali
	Complete nucleotid
	Total DNA sequence
	S. aureofaciens DN
	Human adenosine A1
	Ty lactone synthase
	PRV glycoprotein g
	Single chain antiq
	CDNA encoding huma
	Nucleotide sequenc
	Human colon cancer

PT sequences -
XX
PS Disclosure; Page 85; 207pp; English.
XX
CC This invention relates to a method for identifying a nucleotide or
CC polypeptide sequence that may be a drug target, or essential for growth
CC or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092
CC represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium
CC tuberculosis proteins which are potential drug targets. The DNA and
CC protein sequences are used to illustrate the method of the invention. The
CC method involves providing an unknown nucleotide or polypeptide sequences,
CC and comparing it to a number of sequences along with at least one
CC algorithm capable of analysing a functional relationship between
CC nucleotide and polypeptide sequences. The method is useful for
CC characterising the function of nucleic acids and polypeptides that may be
CC useful as a target for a drug or essential for the growth or viability of
CC an organism.
XX
SQ Sequence 810 BP; 143 A; 249 C; 281 G; 137 T; 0 other;

Query Match	3.48;	Score 47.4;	DB 22;	Length 810;
Best Local Similarity	54.28;	Pred. No. 0.0035;		
Matches 96; Conservative	0;	Mismatches 81;	Indels 0;	Gaps 0;

```

QY 1070 cgttcgagcgcactccgctcccgctggaagtcgcggtttgaggcgcgactactgcttga 1129
    ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 56 cgtcgatcgcgcttcacatcgacacggtgtagccacagagagaagggcgccccaagctggtctca 115

```

QY 1130 cggccttcgcttgctgctctgtgclacagcggtgacggcggtgcgagctgagccgac 1189
| | | | | | | | | | | | | | | | | | | | | |
Db 116 ccgggttcgacccgctgctgctgctgattcagcgcatcacccgacccgctgcccgaagaagccc 175

QY	1190	tgcggcgcgacatcgcaatcccccaagcctlttgtagtgtagcttgccggttgg	124
Db	176	cgctgctcgaaactcgacgctgcacaaacgaggaagaactggtgcagcttggccggtcggtg	232

RESULT	2
AAQ78915	
ID	AAQ78915 standard; DNA; 2232 BP.

AC	AAQ78915;
XX	
DT	03-JUL-1995 (first entry)
DT	

DE *Mycobacterium bovis* PS5 operon.
XX
KW Isoniazid; Isonicotinic acid hydrazide; INH; Inha gene; ps5 gene

05 *Mycobacterium bovis*.

Feature	Key	Location/Qualifiers
FT	CDS	494..1237
FT		/*tag= a
FT		/label= ORF1
FT	CDS	1256..2065
FT		/*tag= b
FT		/label= ORF2
FT		/note= "Inha gene"

PN W09426312-A.

PD 24-NOV-1994.

AA 12-MAY-1994; 94WO-US05344.
PF

PR 13-MAY-1993; 93NZ-0247620.

PR	14-MAY-1993;	93US-0062409.
PR	31-MAR-1994;	94US-0221742.

PA (BANE/) BANERJEE A.

PA (COLL/) COLLINS D M.
PA (DLIS/) DE LISLE G W.
PA (JACO/) JACOBS W R.
PA (WILS/) WILSON T M.
PA (AGRE-) AGRESEARCH.
PA (COLL/) COLLINS D.
PA (YESH) UNIV YESHIVA EINSTEIN COLLEGE

PI Banerjee A, Collins DM, De Lisle GW, Jacobs WR;
PI Wilson TM, Collins D;

DR WPI; 1995-006366/01.
DR P-PSDB; AAR66291; AAR66292.

PT Gene target for isonicotinic acid hydrazide - used to develop
PT prods for diagnosis, treatment, prevention and studies involving
PT mycobacterial infections

PS Disclosure; Fig. 9A-9D; 76pp; English.

CC A cosmid library from *M. bovis* G4/100 was prepared in shuttle
CC vector pYUB18 and transformed into *Mycobacterium smegmatis* ms2155
CC The smallest plasmid obtained which conferred resistance to Inha,
CC the target of action for Isoniazid, was designated pS5. The
CC sequence of pS5 was determined (AAQ78915), revealing 2 large open
CC reading frames, the Inha gene comprising ORF2.

Sequence 2232 BP; 411 A; 696 C; 752 G; 373 T; 0 other;

Query Match	3.4%;	Score 47.4;	DB 16;	Length 2232;
Best Local Similarity	54.28;	Pred. No. 0.0049;		
Matches 96; Conservative	0;	Mismatches 81;	Indels 0;	Gaps 0;

QY 1070 cgttagcgcctactcgtcccgctggaagtcgcggtttagggcgcgactactgagctga 1129
||| || ||| | | ||| ||| | | ||| ||| | ||| ||| | |
Db 1311 cgtcgatcgcttcaatcgcacgcggttagcccaaggagcagggccccaagctgltctca 1370

QY 1130 cgcacctcgcctgtgctgcgtgtcaacgcgcgttgacgcgcgtgtgcgcagatgaaccgac 1189
| | |||| | |||| | | | | | | | |
Db 1371 ccgcggttcgacccgcgcctcgctgatctcagcgcacataccgcacccgcgcgtgcgcgcacaaggccc 1430

QY 1190 tgcgagcgagatcgcaatccccaagccttggtagtgcagcttggccggttgg 1246
 |||||
 Db 1431 cgctgctcgaactcgacgtgcgcaaaacgagagcaccctggccagcttggccgagcggg 1487
 |||||

RESULT	3
AAQ75519	
ID	AAQ75519 standard; DNA; 2232 BP.

AAQ75519;

AA	03-JUL-1995	(first entry)
DT		

DE Mycobacterium bovis PS5 operon.

KM Isonazid; isonicotinic acid hydrazide; INH; inhA gene; vaccine,
KM P55; ss.

OS Mycobacterium bovis G4/100.

AA	Key	Location/Qualifiers
FH		
AA		

```

CDS      494..1234
FT
/*tag= a

```

E1	/LAD= ORF1
FT	1256..2062
CDS	

E1	/lag = 0
ET	/label = ORF2

	gene	/note	XX
	innu	-alou	XX
	nnu	-alou	XX

FN MO3420703-A
XX

Page 2

PD 24-NOV-1994.
XX
PF 13-MAY-1994; 94WO-US05398.
XX
PR 13-MAY-1993; 93NZ-0247620.
PR 14-MAY-1993; 93US-0062409.
PR 31-MAR-1994; 94US-0221742.
XX
PA (AGRE-) AGRESEARCH.
PA (BANE/) BANERJEE A.
PA (COL/) COLLINS D.
PA (JACO/) JACOBS W R.
PA (YESH) UNIV YESHIVA EINSTEIN COLLEGE.
PA (WILS/) WILSON T M.
XX
PI Banerjee A, Collins D, De LISLE GW, Jacobs WR, Wilson TM;
XX
DR WPI: 1995-006691/01.
DR P-PSDB: AAR63899; AAR63900.
XX
PT Polynucleotide(s) determining mycobacterial resistance to
PT isoniazid - useful in diagnosis, treatment and prevention of
PT mycobacterial infection, e.g. tuberculosis.
XX
PS Disclosure; Page 44-48; 104pp; English.
XX
CC A cosmid library of Mycobacterium bovis G4/100 was made in vector
CC pYU818 and transferred into Mycobacterium smegmatis mc2155. The
CC smallest plasmid obtained which conferred an isoniazid resistance
CC phenotype on M. smegmatis was designated pS5. Sequencing of pS5
CC revealed 2 open reading frames, ORF2 corresponding to the *inhA*
CC gene. Mutants of this gene have been used in recombinant vaccines.
XX
SQ Sequence 2232 BP; 411 A; 696 C; 751 G; 374 T; 0 other;

Query Match 3.4%; Score 47.4; DB 16; Length 2232;
Best Local Similarity 54.2%; Pred. No. 0.0049;
.Matches 96; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

OY 1070 cgttgagcgctactccgctccgctggaagtcggttgaggcggaactagtgttga 1129
||| ||||| | | | | | | | | | | | | | | | | |
Db 1311 cgtcgatcgcttccatcgcaagggttagcccaaggagcgcccaactgtgtctca 1370
OY 1130 cggccttcgcttgctgtgtgtgtaacggcggtgacggcggtgcggaagtgagccgac 1189
| | | | | | | | | | | | | | | | | | | | | |
Db 1371 ccgggttcgaccggtgcggtgattcagcgatcacgccgacctgcccgaagggccc 1430
OY 1190 tgcggcgcgatcgcaatccccaagcctttgttgatgtagcttgccggtttg 1246
|| | | | | | | | | | | | | | | | | | | | |
Db 1431 cgctgctcgaaactcgacgtgcaaaaagagagacacttgccggttgccggccggg 1487

RESULT 4
AAT99224
ID AAT99224 standard; DNA; 2232 BP.

XX AAT99224;

DT 01-APR-1998 (first entry)

XX M. bovis pS5 operon.

DE INH resistance; *inhA* gene; isonicotinic acid hydrazide; mycobacteria;

KW mycolic acid biosynthesis; antibiotic susceptibility; infection therapy;

KW INH type antibiotic; M. tuberculosis complex; isoniazid; tuberculosis;

XX antibiotic resistant strain; pS5 operon; ss.

OS Mycobacterium bovis.

XX Key Location/Qualifiers
FH CDS 494..1237
FT /*tag= a

FT CDS 1256..2065
FT /*tag= b
XX
PN US5686590-A.
XX
PD 11-NOV-1997.
XX
PF 12-MAY-1994; 94US-0241766.
XX
PR 12-MAY-1994; 94US-0241766.
PR 14-MAY-1993; 93US-0062409.
XX
PA (AGRE-) AGRESEARCH NEW ZEALAND PASTORAL AGRIC.
XX
PI Banerjee A, Collins DM, De Lisle GW, Jacobs WR;
PI Wilson TM;
XX
DR WPI: 1997-558202/51.
DR P-PSDB: AAW40806, AAW40809.
XX
PT Nucleic acids encoding mycobacterial *InhA* proteins - associated with
PT isoniazid susceptibility and resistance
XX
PS Disclosure; Column 39-44; 55pp; English.
XX
CC This sequence represents the coding sequence for a mycobacterial *InhA*
CC protein. The encoded protein is involved in mycolic acid biosynthesis.
CC The nucleic acids are useful in assessing the susceptibility of various
CC strains of the M. tuberculosis complex to isoniazid (INH) type
CC antibiotics, as decoys and as antisense oligonucleotides to prevent the
CC expression of polypeptides associated with isoniazid (also referred to as
CC isonicotinic acid hydrazide) resistance, and for the expression of the
CC proteins, which may also have use in immunoassays for the detection of
CC INH resistant strains, in the determination of whether an INH type
CC antibiotic may be effective against tuberculosis, and in the treatment of
CC individuals for infection with these strains.
XX
SQ Sequence 2232 BP; 411 A; 696 C; 751 G; 374 T; 0 other;

Query Match 3.4%; Score 47.4; DB 18; Length 2232;
Best Local Similarity 54.2%; Pred. No. 0.0049;
Matches 96; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

OY 1070 cgttgagcgctactccgctccgctggaagtcggttgaggcggaactagtgttga 1129
||| ||||| | | | | | | | | | | | | | | | | |
Db 1311 cgtcgatcgcttccatcgcaagggttagcccaaggagcgcccaactgtgtctca 1370
OY 1130 cggccttcgcttgctgtgtgtgtaacggcggtgacggcggtgcggaagtgagccgac 1189
| | | | | | | | | | | | | | | | | | | | | |
Db 1371 ccgggttcgaccggtgcggtgattcagcgatcacgccgacctgcccgaagggccc 1430
OY 1190 tgcggcgcgatcgcaatccccaagcctttgttgatgtagcttgccggtttg 1246
|| | | | | | | | | | | | | | | | | | | | |
Db 1431 cgctgctcgaaactcgacgtgcaaaaagagagacacttgccggttgccggccggg 1487

RESULT 5
AAH02108
ID AAH02108 standard; DNA; 2609 BP.

XX AAH02108;

DT 24-JUL-2001 (first entry)

DE Mycobacterium tuberculosis nucleotide sequence SEQ ID NO:2101.

KW Species specific; genus specific; family specific; probe; detection;

KW identification; algal; archaeal; bacterial; fungal; parasitical;

KW microorganism; diagnosis; translation elongation factor Tu; toxin;

KW translation elongation factor G; RecA recombinase; resistance;

KW catalytic subunit of proton-translocating ATPase; antimicrobial;

KW vaccine; primer; ds.


```
XX OS Mycobacterium tuberculosis.
XX PN MO200123604-A2.
XX PD 05-APR-2001.
XX PF 28-SEP-2000; 2000MO-CA01150.
XX PR 28-SEP-1999; 99CA-2283458.
XX PR 19-MAY-2000; 2000CA-2307010.
XX PA (INFE-) INFECTIO DIAGNOSTIC (IDI) INC.
XX PI Bergeron MG, Boissinot M, Huletsky A, Menard C, Ouellette M;
XX PI Picard FJ, Roy PH;
XX DR WPI; 2001-245006/25.
XX PT Nucleic acid sequences are used to generate universal probes and
XX PT primers which can be used to identify and detect the presence of algal,
XX PT archaeal, bacterial, fungal and parasitical species in a test sample -
XX PS Disclosure; Page 1488-1489; 1580pp; English.
XX CC The present invention describes a method for generating a repertory of
XX CC nucleic acids of tuf, fus, atpD and/or recA genes from which probes
XX CC and/or primers are derived. The method comprises amplifying the nucleic
XX CC acids of determined algal, archaeal, bacterial, fungal and parasitical
XX CC species with a combination of defined primer pairs. The method can be
XX CC used for producing probes and/or primers for detecting one or more
XX CC related microorganisms e.g. algae, archaea, bacteria, fungi and
XX CC parasites, for universal detection and for specific and ubiquitous
XX CC detection and identification of an algal, archaeal, bacterial, fungal
XX CC and parasitical species, genus, family and group. A nucleic acid (I)
XX CC obtained using the method of the invention can be used for the universal
XX CC detection of any bacterium, fungus or parasite in a sample and for the
XX CC detection of at least one antimicrobial agent resistance gene or at
XX CC least one toxin gene. hexA nucleic acids are used for the specific and
XX CC ubiquitous detection and for identification of Streptococcus pneumoniae.
XX CC (I) can be used to design a therapeutic agent which is effective against
XX CC microorganisms. Microbial species or genus or family or phylum or group
XX CC which can be detected include Abiotrophia adiacens, Bordetella sp.,
XX CC Corynebacterium sp., Enterobacteriaceae group, Escherichia coli,
XX CC Mycobacteriaceae family, Pseudomonads group, Streptococcus sp.,
XX CC Neisseria gonorrhoeae and Staphylococcus sp.. Using DNA based tests
XX CC provides faster results than substrate specificity tests as results can
XX CC be determined in an hour and improved accuracy is also achieved.
XX CC AAH00010 to AAH002304 represent nucleotide sequences and primers/probes
XX CC which are given in the exemplification of the present invention.
XX SQ Sequence 2609 BP; 471 A; 820 C; 875 G; 443 T; 0 other;

Query Match          3.4%; Score 47.4; DB 22; Length 2609;
Best Local Similarity 54.2%; Pred. No. 0.0052;
Matches 96; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 1070 cgttgagcgctactccgctccgctggaagtcgcggttgaggcggaactactggcttga 1129
    ||| ||| ||| | | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1855 cgtcgatcgcttcaacatcgacgaggttagccagagagagggcgccacgtgtgtctca 1914
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1130 cggccttcgctgtgctgtgtgtacagcggtgacggcggtgagcgaggtgagcgac 1189
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1915 ccgggttcgacccgctgcgctgattcagcgcatcacccgacccgctgcgcgcaagccc 1974
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1190 tgcggcgcgatcgcaatccccaagccttltgtatgtagctagcttgccggttgg 1246
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1975 cgctgtcgaactcgacgtgtgcaaaacagagagacactggcagcttgccgcgcg 2031
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 6
AAQ75518

```
ID AAQ75518 standard; DNA; 3120 BP.
XX AC AAQ75518;
XX DT 03-JUL-1995 (first entry)
XX DE M. tuberculosis inhA operon.
XX KW Isoniazid; isonicotinic acid hydrazide; INH; inhA gene; vaccine; ss.
XX OS Mycobacterium tuberculosis.
XX PN WO9426765-A.
XX PD 24-NOV-1994.
XX PF 13-MAY-1994; 94WO-US05398.
XX PR 13-MAY-1993; 93NZ-0247620.
XX PR 14-MAY-1993; 93US-0062409.
XX PR 31-MAR-1994; 94US-0221742.
XX PA (AGRE-) AGRESEARCH.
XX PA (BANE/) BANERJEE A.
XX PA (COLL/) COLLINS D.
XX PA (JACO/) JACOBS W R.
XX PA (YESH ) UNIV YESHIVA EINHSTEIN COLLEGE.
XX PA (WILS/) WILSON T M.
XX PI Banerjee A, Collins D, De LISLE GW, Jacobs WR, Wilson TM;
XX DR WPI; 1995-006691/01.
XX DR P-PSDB; AAR66901.
XX PT Polynucleotide(s) determining mycobacterial resistance to
XX PT isoniazid - useful in diagnosis, treatment and prevention of
XX PT mycobacterial infection, e.g. tuberculosis.
XX PS Disclosure; Page 42-44; 104pp; English.
XX CC The gene from Mycobacterium tuberculosis encoding inhA (AAR66901),
XX CC the target of action for Isoniazid, was identified, isolated,
XX CC cloned and sequenced (AAQ75518). Mutant inhA genes have been used
XX CC for recombinant vaccine development.
XX SQ Sequence 3120 BP; 566 A; 995 C; 1037 G; 522 T; 0 other;

Query Match          3.4%; Score 47.4; DB 16; Length 3120;
Best Local Similarity 54.2%; Pred. No. 0.0056;
Matches 96; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 1070 cgttgagcgctactccgctccgctggaagtcgcggttgaggcggaactactggcttga 1129
    ||| ||| ||| | | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1041 cgtcgatcgcttcaacatcgacgaggttagccagagagagggcgccacgtgtgtctca 1100
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1130 cggccttcgctgtgctgtgtgtacagcggtgacggcggtgagcgaggtgagcgac 1189
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1101 ccgggttcgacccgctgcgctgattcagcgcatcacccgacccgctgcgcgcaagccc 1160
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1190 tgcggcgcgatcgcaatccccaagccttltgtatgtagctagcttgccggttgg 1246
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1161 cgctgtcgaactcgacgtgtgcaaaacagagagacactggcagcttgccgcgcg 1217
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 7
AAQ78914
ID AAQ78914 standard; DNA; 3120 BP.
XX AC AAQ78914;
XX DT 03-JUL-1995 (first entry)
XX

```

DE M. tuberculosis inhA gene.
XX
KW Isoniazid; isonicotinic acid hydrazide; INH; inhA gene;
XX vaccine; mycobacteria; ds.
XX
OS Mycobacterium tuberculosis.
XX
PN W09426312-A.
XX
PD 24-NOV-1994.
XX
PF 12-MAY-1994; 94WO-US05344.
XX
PR 13-MAY-1993; 93NZ-0247620.
PR 14-MAY-1993; 93US-0062409.
PR 31-MAR-1994; 94US-0221742.
XX
PA (BANE/) BANERJEE A.
PA (COLL/) COLLINS D M.
PA (DLIS/) DE LISLE G W.
PA (JACO/) JACOBS W R.
PA (WILS/) WILSON T M.
PA (AGRE-) AGRESEARCH.
PA (COLL/) COLLINS D.
PA (YESH ) UNITV YESHIVA EINSTEIN COLLEGE.
XX
PI Banerjee A, Collins DM, De Lisle GW, Jacobs WR;
PI Wilson TM, Collins D;
XX
DR WPI: 1995-006366/01.
DR P-PSDB; AAR66290.
XX
PT .Gene target for isonicotinic acid hydrazide - used to develop
PT prods for diagnosis, treatment, prevention and studies involving
PT mycobacterial infections
XX
PS Disclosure; Fig. 8A1-8C2; 76pp; English.
XX
CC .
CC Genes from Mycobacterium smegmatis, Mycobacterium tuberculosis and
CC Mycobacterium bovis that encode inhA, the target of action for
CC isoniazid, were identified, isolated, and cloned. Sequences of
CC the 3 genes are given in AAQ78913-15, and encoded amino acids in
CC AAR66289-91. Mutant genes have been used in recombinant vaccine
CC development.
XX
SO Sequence 3120 BP; 562 A; 1006 C; 1040 G; 512 T; 0 other;

Query Match 3.4%; Score 47.4; DB 16; Length 3120;
Best Local Similarity 54.2%; Pred. No. 0.0056;
Matches 96; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 1070 cgttagcgactactcgcgtccgctggaagtcgcggttgaggcggaactactggcttga 1129
   ||| ||||| | | ||| | ||| | | ||||| | ||||| |
Db 1041 cgtcgcgcgcttcacatcgacgcgaggttagccacgagcagggcgccacgtgtgtca 1100
   ||| ||||| | | ||| | ||| | | ||||| | ||||| |
QY 1130 cggccttcgcttgctgtcgtgtgtaacggcggtgaacggcggtgcggcaggtgacggac 1189
   || ||||| ||||| ||||| | | ||| | ||||| | ||||| |
Db 1101 ccgggttcgaccgctgcgctgattcagcgcatcacgcgacggctgcggcaagaagccc 1160
   || ||||| ||||| ||||| | | ||| | ||||| | ||||| |
QY 1190 tgcggcgcgatcgcaatccccaaagcctttgttgatgtagcttgccggttgg 1246
   || ||||| | | ||| | ||| | | ||||| ||||| ||||| ||
Db 1161 cgctgctcgaactcgacgtgcgcaaaacgagagcagcacttgccgacttgccgcgagg 1217
   || ||||| | | ||| | ||| | | ||||| ||||| ||||| ||

RESULT 8
AAT99223
ID AAT99223 standard; DNA; 3120 BP.
XX
AC AAT99223;
XX
DT 01-APR-1998 (first entry)
XX

```

DE	M. tuberculosis InhA gene.
XX	
KW	INH resistance; InhA gene; isonicotinic acid hydrazide; mycobacteria;
KW	mycolic acid biosynthesis; antibiotic susceptibility; infection therapy;
KW	INH type antibiotic; M. tuberculosis complex; isoniazid; tuberculosis;
KW	antibiotic resistant strain; ds.
XX	
OS	Mycobacterium tuberculosis.
XX	
PN	US5686590-A.
XX	
PD	11-NOV-1997.
XX	
PF	12-MAY-1994; 94US-0241766.
XX	
PR	12-MAY-1994; 94US-0241766.
PR	14-MAY-1993; 93US-0062409.
XX	
PA	(AGRE-) AGRESEARCH NEW ZEALAND PASTORAL AGRIC.
XX	
PI	Banerjee A, Collins DM, De Lisle GW, Jacobs WR;
PI	Wilson TM;
XX	
DR	WPI; 1997-558202/51.
XX	
PT	Nucleic acids encoding mycobacterial InhA proteins - associated with
PT	isoniazid susceptibility and resistance
XX	
PS	Disclosure; Column 37-40; 55pp; English.
XX	
CC	This sequence represents the coding sequence for a mycobacterial InhA
CC	protein. The encoded protein is involved in mycolic acid biosynthesis.
CC	The nucleic acids are useful in assessing the susceptibility of various
CC	strains of the M. tuberculosis complex to isoniazid (INH) type
CC	antibiotics, as decoys and as antisense oligonucleotides to prevent the
CC	expression of polypeptides associated with isoniazid (also referred to as
CC	isonicotinic acid hydrazide) resistance, and for the expression of the
CC	proteins, which may also have use in immunoassays for the detection of
CC	INH resistant strains, in the determination of whether an INH type
CC	antibiotic may be effective against tuberculosis, and in the treatment of
CC	individuals for infection with these strains.
XX	
SQ	Sequence 3120 BP; 566 A; 995 C; 1037 G; 522 T; 0 other;
Query Match 3.4%; Score 47.4; DB 18; Length 3120;	
Best local Similarity 54.2%; Pred. No. 0.0056;	
Matches 96; Conservative 0; Mismatches 81; Indels 0; Gaps 0;	
QY	1070 cgttgagcgcctactccgtccgcgtggaagtcgcggttgaggcggaactactggcttga 1129
	111 111111 1 1 11 1 1 11 1 1 11 11111 1 111 1 1
Db	1041 cgtcgcagtcgcgtttcacatcgcacgaggtagccacgagcagggcccccagctgtgtca 1100
QY	1130 cgcgcctcgccttgctgtcgtcgtgtaacgagcggtgaacgcggtcggcaggtgagcgaac 1189
	1 1 1111 111111 1111 1 1 11 1 1 11 1 1 11 1 1 1
Db	1101 ccgggttcgacgcggtgcggtggtcagtcagcgcataccgacgcggtgcgcgcaagccc 1160
QY	1190 tgcggcgcgcatcgcaatcccccaagccttttggtgatgtggtggtggccggttgg 1246
	1 1 11111 1 1 1 111 1 1111 1111111111 1 1
Db	1161 cgtcgtctcgaactcgacgtgcaaaaagagagagcactgtgcccagcttggccgcccggg 1217
RESULT 9	
AAV20433	
ID	AAV20433 standard; DNA; 3120 BP.
XX	
AC	AAV20433;
XX	
DT	15-JUN-1998 (first entry)
XX	
DE	Mycobacterium tuberculosis InhA gene.
XX	
KM	Mycobacterium tuberculosis: InhA gene; enzyme crystal; herbicide;

```
KW X-ray crystallography; three dimensional structure; crystallised; ss.
XX
OS Mycobacterium tuberculosis.
XX
PN US5702935-A.
XX
PD 30-DEC-1997.
XX
PF 28-APR-1994; 94US-0234011.
XX
PR 28-APR-1994; 94US-0234011.
XX
PA (YESH ) UNIV YESHIVA EINSTEIN COLLEGE.
XX
PI Blanchard J, Jacobs WR, Sacchetti J;
XX
DR WPI; 1998-076420/07.
XX
XX
PT Crystalline forms of Inha enzyme - to help in developing
PT anti-tuberculosis agents or herbicides
XX
XX
PS Example 1; Column 15-18; 22pp; English.
XX
CC The present sequence represents the Inha gene from Mycobacterium
CC tuberculosis. The present invention describes: (1) a crystallised
CC Inha enzyme in the form of a plate having the space group C2 and the
CC unit cell constants a = 101.1 Angstrom , b = 83.4 Angstrom , c = 192.9
CC Angstrom , beta = 95 deg. and alpha = gamma = 90 deg.; (2) a
CC crystallised Inha enzyme in the form of a parallelepiped having the
CC space group P21 and the unit cell constants a = 69.0 Angstrom , b =
CC 116.0 Angstrom , c = 65.0 Angstrom , beta = 97.8 deg. and alpha =
CC gamma = 90 deg.; (3) a crystallised Inha enzyme in the form of a
CC hexagon having the space group P622 and the unit cell constants a =
CC b = 100.14 Angstrom , c = 140.4 Angstrom , alpha = beta = 90 deg. and
CC gamma = 120 deg. . A knowledge of the crystal structure of Inha (which
CC is the target of isoniazid in Mycobacterium tuberculosis) will help in
CC the development of bactericides or herbicides capable of inhibiting
CC Inha activity.
XX
SQ Sequence 3120 BP; 566 A; 995 C; 1037 G; 522 T; 0 other;

Query Match 3.4%; Score 47.4; DB 19; Length 3120;
Best Local Similarity 54.2%; Pred. No. 0.0056;
Matches 96; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 1070 cgttgagcgctactccgtcccgctggaagtcggttgaggcggaactactggcttga 1129
   ||| ||| ||| | | | | | | | | | | | | | | | | | | | | | | | |
DB 1041 cgtcgatcgcttccatcgcgcggtgagccagagcagggcgcccaagtgtgtctca 1100

QY 1130 cggcctcgcttggtgtgtgtgtgtacagcggtgacggcggtgagcaggtgagccgac 1189
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1101 ccgggttcgaccggtgtgctgtgattcagcgcataccgaccggtgtccgcaaggccc 1160

QY 1190 tgcggcgcatcgcaatccccaagccttltgtgtgtgtgtgtgtgtgtgtgtgtgtg 1246
   ||| ||| ||| | | | | | | | | | | | | | | | | | | | | | | | |
DB 1161 cgtcgctcgaaactcgacgtgcaaaagagagagcactgtgcccagcttgccgcccgg 1217

RESULT 10
AAV21816
ID AAV21816 standard; DNA; 3120 BP.
XX
AC AAV21816;
XX
DT 14-MAY-1999 (first entry)
XX
DE M. tuberculosis Inha gene.
XX
KW Inha; enzyme; inhibitor; infection; therapy; ss.
XX
OS Mycobacterium tuberculosis.
XX
```

```
PN US5882878-A.
XX
PD 16-MAR-1999.
XX
PF 21-AUG-1996; 96US-0701062.
XX
PR 16-SEP-1994; 94US-0307376.
XX
PR 28-APR-1994; 94US-0234011.
XX
PR 16-JUN-1995; 95US-0491146.
XX
PR 21-AUG-1996; 96US-0701062.
XX
PA (YESH ) UNIV YESHIVA EINSTEIN COLLEGE.
XX
PI Sacchetti J;
XX
DR WPI; 1999-214062/18.
XX
XX
PT Identifying inhibitors of Inha enzyme - by identifying compound
PT that fits the catalytic active site of the crystallised enzyme,
PT useful for treating tuberculosis
XX
XX
PS Disclosure; Column 3-8; 9pp; English.
XX
CC This sequence represents the Mycobacterium tuberculosis Inha coding
CC sequence. The invention relates to a method for identifying inhibitors of
CC Inha enzyme comprises identifying compounds which fit the catalytic
CC active site of the crystallized enzyme. Inhibitors of the activity of the
CC Inha enzyme can be used for treating Mycobacteria tuberculosis infection.
XX
SQ Sequence 3120 BP; 566 A; 995 C; 1037 G; 522 T; 0 other;

Query Match 3.4%; Score 47.4; DB 20; Length 3120;
Best Local Similarity 54.2%; Pred. No. 0.0056;
Matches 96; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 1070 cgttgagcgctactccgtcccgctggaagtcggttgaggcggaactactggcttga 1129
   ||| ||| ||| | | | | | | | | | | | | | | | | | | | | | | | |
DB 1041 cgtcgatcgcttccatcgcgcggtgagccagagcagggcgcccaagtgtgtctca 1100

QY 1130 cggcctcgcttggtgtgtgtgtgtacagcggtgacggcggtgagcaggtgagccgac 1189
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1101 ccgggttcgaccggtgtgctgtgattcagcgcataccgaccggtgtccgcaaggccc 1160

QY 1190 tgcggcgcatcgcaatccccaagccttltgtgtgtgtgtgtgtgtgtgtgtgtgtg 1246
   ||| ||| ||| | | | | | | | | | | | | | | | | | | | | | | | |
DB 1161 cgtcgctcgaaactcgacgtgcaaaagagagagcactgtgcccagcttgccgcccgg 1217

RESULT 11
AAV68212
ID AAV68212 standard; DNA; 3120 BP.
XX
AC AAV68212;
XX
DT 25-JAN-1999 (first entry)
XX
DE Inha gene nucleotide sequence.
XX
DE Inha enzyme; Mycobacterium tuberculosis strain H37 RV; bactericide;
KW enoyl reductase activity; mycolic acid production; herbicide; ss.
XX
OS Mycobacterium tuberculosis.
XX
XX
PN US5837480-A.
XX
PD 17-NOV-1998.
XX
PF 21-AUG-1996; 96US-0700306.
XX
PR 28-APR-1994; 94US-0234011.
XX
PR 21-AUG-1996; 96US-0700306.
XX
```

PA (YESH) UNIV YESHIVA EINSTEIN COLLEGE.
XX
PI Blanchard J, Jacobs WR, Sacchetti J;
XX
DR WPI; 1999-023444/02.
XX
PT Use of Inha enzyme in assays - for determining whether compounds
PT inhibit Inha-mediated enoyl reductase activity, for identifying
PT bacteriocides or herbicides
XX
PS Disclosure; Columns 3-8; 23pp; English.
XX
CC The present sequence represents the inha gene of Mycobacterium
CC tuberculosis strain H37 RV. The inha enzyme is used to identify
CC compounds which inhibit inha-mediated enoyl reductase activity.
CC Compounds which inhibit the activity of Inha enzyme inhibit
CC its ability to produce mycolic acids which are essential for cell
CC survival. Such compounds can be used to treat bacterial infection,
CC particularly mycobacterial infection such as Mycobacterium tuberculosis
CC infection. In addition such compounds can be used as herbicides.
XX
SQ Sequence 3120 BP; 566 A; 995 C; 1037 G; 522 T; 0 other;

Query Match 3.4%; Score 47.4; DB 20; Length 3120;
Best Local Similarity 54.2%; Pred. No. 0.0056;
Matches 96; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 1070 cgtgagcgctactccgtcccgctggaagtcgagtgtagggcgactagcttga 1129
Db 1041 cgtcgatcgcttccatccgacgagtgtagccagagagcgccagctgtgtca 1100
QY .1130 cggccttcgcttgctgctgctgacgagcggtgacgagcggtgagtgagcgac 1189
Db 1101 cggggttcgacgagctgctgctgacgacatcacgacgagcggtgagcgagcgcc 1160
QY 1190 tggcgcgagtcgcaatcccaagccttggtagtgtagcttgccggttg 1246
Db 1161 cgtgctcgactcgacgtgcaaaagagagagcactgagcgcttgccgagcgcg 1217

RESULT 12
AAA10594/c
ID AAA10594 standard; DNA: 10732 BP.
XX
AC AAA10594;
XX
DT 29-JUN-2000 (first entry)
XX
DE Gene encoding a subunit of cellulose synthase.
XX
KW Cellulose synthase; cellulose production; increase yield; ds.
XX
OS Vigna angularis.
XX
PN JP2000060568-A.
XX
PD 29-FEB-2000.
XX
PE 26-AUG-1998; 98JP-0239998.
XX
PR 26-AUG-1998; 98JP-0239998.
XX
PA (MIZU/) MIZUNO K.
PA (OJIP) OJI PAPER CO.
DR WPI; 2000-342371/30.
DR P-PSDB; AAY85179.
XX
PT A gene encoding a cellulose synthetic equipment - for the improvement
PT in the amount of cellulose synthesised in a plant body
XX
PS Claim 2; Page 14-21; 32pp; Japanese.

XX
CC This sequence represents a gene encoding a subunit of the cellulose
CC synthase complex of Vigna angularis. The invention relates to subunits of
CC cellulose synthetic equipment, that can be used to increase the amount of
CC cellulose synthesised by a plant. The proteins and genes encoding them
CC can also be used to improve the properties of the cellulose being
CC produced by a plant.
XX
SQ Sequence 10732 BP; 3149 A; 1212 C; 2074 G; 2046 T; 2251 other;

Query Match 3.2%; Score 44.4; DB 21; Length 10732;
Best Local Similarity 13.2%; Pred. No. 0.055;
Matches 178; Conservative 509; Mismatches 659; Indels 5; Gaps 4;

QY 26 ttgccattaccaaccacagtgaggccacagcttctgcatcgctgttgca 85
Db 10715 YTSRSRCYCTSRKCYTAYSRCSYRSRSTYTBTTYSNSTYDANSTCYTSCSD 10656
QY 86 gctcgagcttgaggcgctccagcagctgtgttgctgagcagtggtgct 145
Db 10655 RCSRTRBRNSTBRSRCYCTBRCSNSTYSNSTRCSTYSSTSRYNCTSRYSSTTTAKNC 10596
QY 146 tctgctgtgt-cgttcaagctcgctcgttgtgcccagtlccgcccagggtg 204
Db 10595 STYSCYTYDASRTBSRBSRYSDYYSSTBNSCTCYTYDAYSYDAYDATYDANSTNS 10536
QY 205 gggctagccgagctcgagcctattggccctgctcctcgctgacagatatcgatcg 264
Db 10535 TYDATBYDANSTDNCTBYATTTBTBYSSSTDYSTSRYTTTAKTTSRNCNRNSTYDA 10476
QY 265 caagcaaccccatcactgctgtgtgctgcttactgagggcgctgagccctagcaag 324
Db 10475 CNCNRGCTTSRCYSTTYYDACYASTNSTCYTYSRSTSTSDSRYSRGCSSRYSAKY 10416
QY 325 ggaactcaccgctgagcgtcagctttagttgggctagcacaactgagcttact 384
Db 10415 SCNCCSRGSRSTTCCSRGNSTYSNSTSRNSTCCYSDTBSRNSTSDTTTSCCTCYD 10356
QY 385 ttggttlttgcctagcgctcggttctcgcgaatcccgctcgatcgctgttc 444
Db 10355 ASDNCCSTNSTYSNSTSRNCNSTSRNCNSTSCSRDYSNCCCYTSSSTYDASR 10296
QY 445 tcggtcgctgcatcaatcgcttltgtcagtgctacgagcctcaacaaatgatctac 504
Db 10295 CYSCNCCNSTAKYSCCYTYDAYTYSNCSRCYCTBCSRNCYSCCTYSRSTSTYD 10236
QY 505 ggcgttgaagctgagcacttggtgagtcgcaactcggttcgagcttcaactcag 564
Db 10235 ADCSRNSTYRAYRACRNCSTNCSTSDRCSRGYSYDAYDATBNCYYSNSTCCYTSDC 10176
QY 565 gttacagctatctggcaaccccaactgctgctgcttactatctgtgcgaagcagct 624
Db 10175 YTCRCTBSDDYDACNSTCCRCSTYSCTTYDACSTYASTDNCSRSTSTCTAKYSYTB 10116
QY 625 ttltcgaagcagcagcgggtgtgagcgagctgctcccaagctgagcgatcgct 684
Db 10115 STNCYSDSRGSNSTAKY-ASTTYDAYDARCTTDNSTRCYRASTYDASRRCYSSYRAS 10057
QY 685 gcgaagtgtagcagctatgtctgacccaactacagtcgagcggtgctggt 744
Db 10056 RGTBDDCCTBSRSGYSYDASRYANDPBTBAYRARCAYDAYARCNSTSRAYS 9997
QY 745 ttgtcgcatgattttgtctggcggtataggcgtctactgttcaaccccgctcta 804
Db 9996 DSTCYTRCGRSNCSTYSYSTYRASTBT--CYCTBCSRRCYSRYSSSTCNCYSY 9939
QY 805 cccgacccctgagcagctgctattccagctgattggtgagctagtcggtgctc 864
Db 9938 CCYTSRYSSTTNSCTCTYSNSTBTBYSYSTYSRGSRGSDRGNCYNSTNC 9879
QY 865 ttggtgagctgctgagcttagccgctgagcgctgagctgttagcatcttgtgagg 924

Db	9878	YDASTSDTBYSRCCYTYTYSSTDSTDSTDSTDSTNSTBTBDCYTTTBTBSRSTSDSTSTYRCGR	9819
QY	925	cgtgaagacagcagcaacactccgcatcgaatgctctgctgcgctgcagatgatt	984
Db	9818	SDYDATBSDNSTNCCYDASRBTBTBSTNCAICYARCYTBYDARCSRDSTYSRGRAYDANSTSRYS	9759
QY	985	caagatcgccctgtgctgagcatcgccccgcgaatacgcgcttaacctgttatccc	1044
Db	9758	SRYSSTYSSDSTYSAKYCAKSTBTBICYDAYDACYDAYANCYSSDSTYTBYSRRCC	9699
QY	1045	ctctataacagcgcgcttctaacgcg-gttgagcgctactcccgctccgctggaagtcgc	1103
Db	9698	YYDAYSCSRDYDARCYDACYSNSTCYDATBTBSRSTYTYSNCYDATTSRCTBTSTB	9639
QY	1104	ggttgagggcggaactactgagcgttgacgacctgcgttgctgtctgtgtacagcgagt	1163
Db	9638	TBTTSRCAKCTBDSTSTAKNSTSTYSTTRCTBYSSRSRSGYSCSRSRSCYCYTDSST	9579
QY	1164	gaacgcggtgcgagcaggtgagccgactgcgcgcgatcgcaatcccaagcctttggt	1223
Db	9578	CYSTTYSTTYAYSCSTTSRGSYDASRSTSTYSRCTTSTPYSTTDYSDCSDYSTTBMNST	9519
QY	1224	gatgctagacttgccggttttgccaggaatgcttggttcaagtgctctgtttgataccgtgct	1283
Db	9518	YSSDSDCCTBYSSDRCSRDSCTCNCYSCSDSRYSTTYDACTTYDAKTBCTYTSDNCCNST	9459
QY	1284	ctatcgaccggaagccagtaacgctctgtgtgctctgtlattgagcgatcgcgagttctcg	1343
Db	9458	SRCYTNSTYSCSRBTBSRNSTCCCTBTTSRGMCCYDAYDANSTSRDAYYDACYSYDAST	9399
QY	1344	gcagccccaacctccaagaactccctcoa	1374
Db	9398	TBYSYSTCTBYSNSTYDAYSSRYSCTYCTCD	9368

PS	Disclosure; Column 21-24; 55pp; English.
XX	
PT	isoniazid susceptibility and resistance
PT	Nucleic acids encoding mycobacterial InhA proteins - associated with
XX	
DR	P-PSDB; AAW40806.
XX	
PI	Banerjee A, Collins DM, De Lisle GW, Jacobs WR; Wilson TM;
XX	
PA	(AGRE-) AGRESEARCH NEW ZEALAND PASTORAL AGRIC.
XX	
PR	12-MAY-1994; 94US-0241766.
PR	14-MAY-1993; 93US-0062409.
XX	
PE	12-MAY-1994; 94US-0241766.
XX	
PD	11-NOV-1997.
XX	
PN	US5686590-A.
OS	Mycobacterium bovis.
XX	
KW	INH type antibiotic; M. tuberculosis complex; isoniazid; tuberculosis;
KW	antibiotic resistant strain; ss.
XX	
DE	M. bovis INH resistance gene.
XX	
DT	01-APR-1998 (first entry)
XX	
AC	AAT99220;
XX	
ID	AAT99220 standard; DNA; 1723 BP.
RESULT	13
AAT99220	

XX This sequence represents the coding sequence for a mycobacterial InhA
 CC protein. The encoded protein is involved in mycolic acid biosynthesis.
 CC The nucleic acids are useful in assessing the susceptibility of various
 CC strains of the *M. tuberculosis* complex to isoniazid (INH) type
 CC antibiotics, as decoys and as antisense oligonucleotides to prevent the
 CC expression of polypeptides associated with isoniazid (also referred to as
 CC isonicotinic acid hydrazide) resistance, and for the expression of the
 CC proteins, which may also have use in immunoassays for the detection of
 CC INH resistant strains, in the determination of whether an INH type
 CC antibiotic may be effective against tuberculosis, and in the treatment of
 CC individuals for infection with these strains.
 XX
 SQ Sequence 1723 BP; 328 A; 530 C; 580 G; 285 T; 0 other;

Query Match	2.8%	Score 39.6;	DB 18;	Length 1723;
Best Local Similarity	55.1%	Pred. No. 0.58;		
Matches 98;	Conservative 0;	Mismatches 79;	Indels 1;	Gaps 1;
QY 1070	cyttgagcgccactccgctcccgctggaagtcgcggttgaggcgagactactggtctga	1129		
Db 968	cytcgatacgctttcacatcgcacgsggtagccacagagcagcgcgccacagctgtgtctca	1027		
QY 1130	cggccttcgcttgctgtctgtg-gtcacggcggtgacgcgsgtgcgcgcaggtgagccga	1188		
Db 1028	ccgggttcgaccgcgctgcgcgtgtattcacgcataccacgcgagctgcgcgcaaaagccc	1087		
QY 1189	ctgcgcgcgcatcgcaatcccccaagccctttgttgatgtgcagcttgcccggttgg	1246		
Db 1088	ccgctgctcgaaactcgacgtgcgcaaaaacgaggaagcactgtgcgaagcttgcccgccggg	1145		

RESULT	14
AAT99219	
ID	AAT99219 standard; DNA; 1723 BP.
XX	
AC	AAT99219;
XX	
DT	01-APR-1998 (first entry)
XX	
DE	M. tuberculosis INH resistance gene.
XX	
KW	INH resistance; InhA gene; isonicotinic acid hydrazide; mycobacteria;
KW	mycolic acid biosynthesis; antibiotic susceptibility; infection therapy;
KW	INH type antibiotic; M. tuberculosis complex; isoniazid; tuberculosis;
KW	antibiotic resistant strain; ss.
XX	
OS	Mycobacterium tuberculosis.
XX	
PN	US5686590-A.
XX	
PD	11-NOV-1997.
XX	
PF	12-MAY-1994; 94US-0241766.
XX	
PR	12-MAY-1994; 94US-0241766.
PR	14-MAY-1993; 93US-0062409.
XX	
PA	(AGRE-) AGRESEARCH NEW ZEALAND PASTORAL AGRIC.
XX	
PI	Banerjee A, Collins DM, De Lisle GW, Jacobs WR;
PI	Wilson TM;
XX	
XX	WPI: 1997-558202/51.
DR	P-PSDB; AAM40805.
XX	
PT	Nucleic acids encoding mycobacterial InhA proteins - associated with
PT	isoniazid susceptibility and resistance
XX	
PS	Claim 1; Column 19-22; 55pp; English.
XX	
CC	This sequence represents the coding sequence for a mycobacterial inhA

DR AAB07562, AAB07563, AAB07564, AAB07565, AAB07566, AAB07567, AAB07568,
DR AAB07569, AAB07570, AAB07571, AAB07572, AAB07573, AAB07574, AAB07575,
DR AAB07576, AAB07577, AAB07578.

PT New bleomycin gene cluster components useful for peptide and/or
PT polyketide metabolites, especially bleomycin, production and for
PT chemically modifying biological molecules -

Claim 8; Page 97-136; 162pp; English.

CC The present sequence represents the BLM (Bleomycin) gene cluster,
CC containing open reading frames (ORFs) 8-30. The proteins encoded
CC by the gene cluster are useful for producing peptides and/or polyketide
CC metabolites, especially bleomycin or bleomycin analogues. They are
CC also useful for chemically modifying biological molecules to produce
CC branched methyl groups, and for coupling amino acids and fatty
CC acids. They may be reacted with an apo-carrier protein and coenzyme A
CC to produce a holo-carrier protein. The BLM gene cluster or catalytic
CC domains can be used individually or collectively to produce
CC thiazolidine, thiazoline, bithiazoline and bithiazoline-containing
CC microbial metabolites. The BLM gene cluster may also be used to produce
CC sugars.

SQ Sequence 58857 BP; 7256 A; 25139 C; 19353 G; 7109 T; 0 other;

Query Match	2.8%;	Score 39.2;	DB 21;	Length 58857;
-------------	-------	-------------	--------	---------------

```
Matches 95; Conservative 0; Mismatches 93; Indels 0; Gaps 0;
```

QY.	1008	cggccccggaataaccgcctttaacctggttatccctctatcaacagcgcgcttacc	1067
Db	29192	cgaaccaagtgaagaatcgcgcgcgatccgcggttcgaagcccgcgaagtgcgcggaagcgcttcg	29251
QY	1068	ggcgcttagacgcgcctaactccgctccgcgcgtgaagtcgcgcgttgagaggcggaactatcgctt	1127
Db	29252	ggccctggcccgcgctcgcgcgaagcacgcgcggtcgtlccgcacagacggcgcgctggcgcgta	29311
QY	1128	gaacgccttcgcctcgctgtcgtcgtgtacaagcgcggtgaacgcgcgtgcgcaggttagcgcg	1187
Db	29312	cgcgcgtcgcgcgaaccgcgtcgcgcccgcccccgcgcgcgaacgcgcctcgcggaacgcgcgtcg	29371
QY	1188	actgcgcgc	1195
Db	29372	caggcgcgc	29379

Search completed: January 10, 2002, 22:25:47
Job time: 2985 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 10, 2002, 21:32:57 ; Search time 1447.17 Seconds
(without alignments)
16005.058 Million cell updates/sec

Title: US-09-887-038-2
Perfect score: 1404
Sequence: 1 atgactgtctgtgcaactct.....attcagcagcaaaatgttag 1404

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vi:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Query			DB	ID	Description
		Match	Length	DB			
1	1404	100.0	4957	1	SPU62616		U62616 Synechococc
2	326.4	23.2	143051	1	D90911		D90911 Synechocyst
3	49.8	3.5	7218	6	166494		166494 Sequence 14
4	47.4	3.4	810	1	AF106077		AF106077 Mycobacte
5	47.4	3.4	832	1	MTU02492		U02492 Mycobacteri
6	47.4	3.4	1789	1	MBU41388		U41388 Mycobacteri
7	47.4	3.4	2232	6	173546		173546 Sequence 12
8	47.4	3.4	2609	6	AX111368		AX111368 Sequence
9	47.4	3.4	3120	6	126125		I26125 Sequence 11
10	47.4	3.4	3120	6	173545		I73545 Sequence 11
11	47.4	3.4	3120	6	186995		I86995 Sequence 11
12	47.4	3.4	20113	1	AE007022		AE007022 Mycobacte
13	47.4	3.4	38300	1	MTCY277		279701 Mycobacteri
14	44.4	3.2	10732	6	E32986		E32986 Gene encodi
15	42.8	3.0	346294	1	AP002999		AP002999 Mesorhizo
16	41.8	3.0	20021	1	AE004730		AE004730 Pseudomon
17	41.6	3.0	2793	1	AE254803		AF254803 Burkholde
18	41.2	2.9	37890	2	AC092316		AC092316 Homo sapi
19	40.8	2.9	127488	2	AC022095		AC022095 Homo sapi
20	40.4	2.9	45335	1	SCC57A		AL136519 Streptomy
21	39.8	2.8	189333	1	RME603647		AL603647 Rhizobium
22	39.6	2.8	1723	6	I73541		I73541 Sequence 1
23	39.6	2.8	1723	6	I73542		I73542 Sequence 2
24	39.6	2.8	13569	1	AE004772		AE004772 Pseudomon
25	39.6	2.8	70089	2	AC027491		AC027491 Homo sapi
26	39.4	2.8	34766	1	SCC121		AL137166 Streptomy
27	39.4	2.8	137137	9	AC004928		AC004928 Homo sapi
28	39.2	2.8	1246	6	AX164174		AX164174 Sequence
29	39.2	2.8	9316	1	AF077869		AF077869 Streptomy
30	39.2	2.8	77457	1	AF210249		AF210249 Streptomy
31	39.2	2.8	168347	2	AC025336		AC025336 Homo sapi
32	38.8	2.8	5023	3	DMU11383		U11383 Drosophila
33	38.8	2.8	6655	3	DMOVO		X59772 D.melanogas
34	38.8	2.8	12991	1	AE007016		AE007016 Mycobacte
35	38.8	2.8	39150	1	MTCY21B4		Z80108 Mycobacteri
36	38.6	2.7	13283	1	AE005006		AE005006 Halobacte
37	38.6	2.7	172154	2	AC087726		AC087726 Chlamydom
38	38.4	2.7	5186	1	RCAMCPAB		L48927 Rhodobacter
39	38.4	2.7	10757	1	AE004561		AE004561 Pseudomon
40	38.4	2.7	10944	1	AE004771		AE004771 Pseudomon
41	38.2	2.7	1681	1	SC6610A		AL445343 Streptomy
42	38.2	2.7	11477	1	AE005892		AE005892 Caulobact
43	38	2.7	132712	2	AP004009		AP004009 Oryza sat
44	38	2.7	136555	2	AC068279		AC068279 Homo sapi
45	38	2.7	205061	2	AC069341		AC069341 Homo sapi

ALIGNMENTS

RESULT 1
SPU62616
LOCUS SPU62616 4957 bp DNA
DEFINITION Synechococcus PCC7942 putative protein (dcl1) gene, partial cds, and putative proteins (dcl2), (dcl3), (dcl4) and (dcl3) genes, complete cds.
ACCESSION U62616
VERSION U62616.1 GI:1549372
KEYWORDS Synechococcus sp. PCC 7942.
SOURCE Synechococcus sp. PCC 7942.
ORGANISM Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
REFERENCE 1 (bases 1 to 4957)
AUTHORS Bonfil,D.J., Lieman-Hurwitz,J., Ronen-Tarazi,M. and Kaplan,A.
TITLE Genomic region involved in ability of Synechococcus PCC 7942 to grow under low CO2 conditions
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 4957)
AUTHORS Bonfil,D.J., Lieman-Hurwitz,J., Ronen-Tarazi,M. and Kaplan,A.
TITLE Direct Submission

JOURNAL Submitted (01-JUL-1996) Plant Biology, Hebrew University, Givat Ram, Jerusalem 91904, Israel

FEATURES

source

1. .4957
/organism="Synecchococcus sp. PCC 7942"
/strain="PCC 7942"
/db_xref="taxon:1140"
1. .845
/gene="dc11"
<1. .845
/gene="dc11"
/note="similar to the C-terminus of a Synecchocystis PCC6803 putative protein encoded by GenBank Accession Number D64000 (cyanobase N. sl10524)"
/codon_start=3
/transl_table=11
/product="putative protein"
/protein_id="AAB08473.1"
/db_xref="GI:1549373"
/translation="AMIEAIGVNSDDPQFLMVDPFSAKIEDLLQGLDFAYPEAVKVGGLASGLAESAIASLFFQDRQYDGVIGIALSGNVQIATVAQGRPVGPLMHWAAERNLRLQIOTDEDEPIALQALQSVLRDLSPELQSLCVGLACNSFQTVLQPGDELIRNLTGFDPRTGAVAIQDRIRVQRLQHVDAQTAAADLERIQGMCROHATKPAASLLFSCLGKGRKPFYQOANFESQLIQHYLSELPLAGFCNGEIGPIAGSTYLLHGYTSVALILISAKTH"

misc_feature
132. .188
/gene="dc11"
/note="encodes a putative transmembrane region"

misc_feature
147. .158
/gene="dc11"
/note="encodes a glycosaminoglycan attachment site"

*misc_feature
237. .269
/gene="dc11"
/note="encodes a prokaryotic membrane lipoprotein lipid attachment site"

misc_feature
411. .473
/gene="dc11"
/note="encodes a putative transmembrane region"

misc_feature
723. .797
/gene="dc11"
/note="encodes a putative transmembrane region"

gene
complement(1095. .1928)
/gene="d133"
complement(1095. .1928)
/gene="d133"

CDS
complement(1095. .1928)
/gene="d133"
/note="similar to Synecchocystis PCC6803 cyanobase N. sl1895 and to E. coli and H. influenzae ICC proteins, Swiss-Prot Accession Numbers P36650 and P44685 respectively; similar to a 12.8 KDa putative protein in the 16S rRNA gene region of Synecchococcus PCC6301, Swiss-Prot Accession Number P05677"
/codon_start=1
/transl_table=11
/product="putative protein"
/protein_id="AAB08474.1"
/db_xref="GI:1549374"
/translation="MAMNRAFRALLSDLHIGLPQTIWHRHFRHLIECSIPAFEQILTELAIDLDELRLPGDLTOGGERENHQLLOQLROLPPYVVPGNHMDPAASDRDRTGVAEFVELYQDFGFTSDRPYRCTIAPDLDFALNSIAFDEAGHQLHGRVDCQJNWLQOELATSTASQRWVMIHNVLEHLPPDQOQHPLGORYILENAEELRSILEAGSVSYIFGHLHIQATSQOGALWEVTGSLVSPHPYRLGTVPAAEGWQLELESRRVRSIPDQD"

misc_feature
complement(1311. .1322)
/gene="d133"
/note="encodes an endoplasmic reticulum targeting sequence"

gene
1353. .1748
/gene="dc12"
1353. .1748
/gene="dc12"
/codon_start=1
/transl_table=11
/product="putative protein"

/protein_id="AAB08475.1"
/db_xref="GI:1549375"
/translation="MLLLIGOMLOHIYVNHHPALASGSGEFLLOPVELRADSPPMQIMARFIKSDRIQRQIEIRCDRAAIVGAIAAREAKVLIELDKLGHGTGAIAIARCGGHIVVARHIDRIGQALQQLQPLMVFPPLVIG"
complement(1431. .1442)
/gene="d133"
/note="encodes an endoplasmic reticulum targeting sequence"

misc_feature
1683. .1745
/gene="dc12"
/note="encodes a putative transmembrane region"

gene
1998. .2639
/gene="dc13"
1998. .2639
/gene="dc13"
/note="cytoplasmic protein; similar to from Synecchocystis PCC6803 cyanobase N. sl11300, to Escherichia coli YGGH gene product, Swiss-Prot Accession Number P32049, to Haemophilus influenzae YGGH gene product, Swiss-Prot Accession Number P44648 and to a region in the mica/muty gene from Escherichia coli"
/codon_start=1
/transl_table=11
/product="putative protein"
/protein_id="AAB08476.1"
/db_xref="GI:1549376"
/translation="MARVRVRQHVNPISQKROVTTMPDMOQVYADCDRLHLIDIGCARFLLAMATRQPEWNYLGLIEIREPLVDEANAIARERELTNLYHFPNANLDLEPLRLSPTGILQORVSIQFPDPEFKRHKQRRVQPELVQALATAPAGAEVFLQSDVLEVOAEMCEHFAAEPRFRQTCIDMLPENPLPVPTERTAVQNKQLPVYRALFIRQPAD"

misc_feature
2160. .2180
/gene="dc13"
/note="encodes a tyrosine kinase phosphorylation site"

gene
2670. .4073
/gene="dc14"
2670. .4073
/gene="dc14"
/note="inactivation of this gene produced a high-CO2-requiring mutant; similar to Synecchocystis PCC6803 cyanobase N. sl1515 and to Escherichia coli sodium/pantothenate symporter and a multidrug resistance protein, Swiss-Prot Accession Numbers P16256 and P21441 respectively"

CDS
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="AAB08477.1"
/db_xref="GI:1549377"
/translation="MTWQTLTFAHYQPOQMGHSSFLHRLFGSLRAWRASSQLLWSEALGGFLAAVYGSAPFVPSSALGLGLAIAIYAWALLSTIDIDLQATPPIHMLVLLYWGVDALATGLSPVRAALVGLAKITLYLVFADLAARVLRNPRLRSLFSVVVITSLFVSVYGLNQMITYVEELATWVDNRNSVADFTSRVSYLGNPNLLAAYLVPTTAFSAAIGVNRGMLPKLIAIAATGASSLCILITYSRGGWLGFWAMIFVMALLGLVFOPLPAPWRWLFPVVLGGLVAVLVAVLGLLEPLRVLSIFVGREDDSSNFRINWLAVLQMIQDRPWLGIQPGNTAFNLVYVLYQOARFTALSAVSVPLEVAVEGGLGLTAFAMLLLVTAVTAVRQVSRLLRRDRNPQAFWLMAASLAGMLGHGLFDTVLYRPEASTLMWLCIGAIAISFWQPQPSKQLRPEAEHDEKM"

misc_feature
2847. .2918
/gene="dc14"
/note="encodes a putative transmembrane region"

misc_feature
2952. .3035
/gene="dc14"
/note="encodes a putative transmembrane region"

misc_feature
3012. .3074
/note="encodes a putative transmembrane region"

misc_feature
/gene="dc14"
/note="encodes a putative transmembrane region"

misc_feature
3081. .3146
/gene="dc14"
/note="encodes an ATP synthase c subunit signature"

misc_feature
3102. .3158
/gene="dc14"
/note="encodes a putative transmembrane region"

```
misc_feature      3264. .3320
                  /gene="dc14"
                  /note="encodes a putative transmembrane region"
misc_feature      3400. .3455
                  /gene="dc14"
                  /note="encodes a putative transmembrane region"
misc_feature      3492. .3554
                  /gene="dc14"
                  /note="encodes a putative transmembrane region"
misc_feature      3783. .3842
                  /gene="dc14"
                  /note="encodes a putative transmembrane region"
misc_feature      3882. .3941
                  /gene="dc14"
                  /note="encodes a putative transmembrane region"
misc_feature      3972. .4022
                  /gene="dc14"
                  /note="encodes a putative transmembrane region"
BASE COUNT      972 a 1418 c 1451 g 1116 t
ORIGIN
```

```
Query Match      100.0%; Score 1404; DB 1; Length 4957;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 atgactgtctgcaaacctctgaccttttgccattaccaccccaacagtgaggcaaacg 60
Db      2670 ATGACTGTCTGGCAACCTGACTTTTGCCCATTTACCAACCCCAACAGTGGGGCCACAGC 2729

QY      61 agttcttgcatcggtgtttgagcagcctgagagcttgagcgagcctccagccagctgtt 120
Db      .2730 AGTTTCTTGCAATCGGCTGTTGGCAGACCTCGAGCTTGGCGGGCCCTCCAGCCAGCTTTG 2789

QY      121 gtttggtctgagggcactgggtggtcttctgtctgtctgtctacaggttcogctcgtt 180
Db      2790 GTTTGGTCTGAGGCACTGGGTGGCTTCTGTGCTGTCTACGGTTCCGCTCCCTTT 2849

QY      181 gtgcccagttccggccctagaggttgaggctagccgagatcgagcctattggccctgtc 240
Db      2850 GTGCCAGTTCCGCCCTAGGGTTGGGGCTAGCCCGCATCGCGCCCTATTTGGGCCCTGCTC 2909

QY      241 tcgctgacagatatcgatctgcgcgcaagcaaccccatcactgtgctgtgtctctac 300
Db      2910 TCGCTGACAGATATCGATCTGCGGCAAGCAACCCCATTCAC TGCGTGGTGTCTTAC 2969

QY      301 tggggcgtcgatgccccctagcaacgaggactctcaaccgtaacgctgcagcttaagtgg 360
Db      2970 TGGGGCGTCGATGCCCTAGCAACGGGACTCTCACCCCGTACGGCTGCAGCTTTAGTTGGG 3029

QY      361 ctagccaaactgaagccttaacctgtgtttttgcccctagcggctcggttctccgaat 420
Db      3030 CTAGCCAAACTGACGCTCTTACCTGTGTTTGGCCCTTAGCGGCTCGGGTTCTCCGAAT 3089

QY      421 ccccgctcgatcgctgtgtctcgtgcgtcgtgcgtacacatcgctttttgtcagttc 480
Db      3090 CCCCCTCGGATCGCTGTGTCTCGGTGCTGCTGATCACATCGCTTTTGTGCTGTGTC 3149

QY      481 tacggcctcaaccaatgatctaagcggtgaagagctgagcacttggtgatcgcaac 540
Db      3150 TACGGCCTCAACCAATGATCTACGGCGTTGAAGAGCTGGCGACTTGGGTGATCGCAAC 3209

QY      541 tcggttgccgacttcacctcacggygtttacagctatctggygcaaccccaacctgtgct 600
Db      3210 TCGGTTGCCGACTTCACCTCACGGGTTTACAGCTATCTGGGCAACCCCAACCTGTGCT 3269

QY      601 gcttatctgtgcccgaactgcttctctgcagcagcagatcggygtgtgagcgagctg 660
Db      3270 GCTTATCTGTGTGCCGACGACTGCTTTTCTGCAGACGCGATCGGGGTGTGGCGGCTGG 3329

QY      661 ctccccaagctgtgagcagatcgctgcagacagtgcgagcagcttalgtctgacccacc 720
Db      3330 CTCCCAAGCTGTGGGATCGCTGCAGACAGGTGCGAGCAGCTTATGTCTGATCTCTCACC 3389
```

```
QY      721 tacagtcgagtgctgagctgggttttgtcgccatgattttgtctggcgcttattagg 780
Db      3390 TACAGTCGCGGTGGCTGGCTGGGTTTGTGCGCCATGATTTTGTCTGGCGGCTTATAGGG 3449

QY      781 ctctactggttcaaccccgctaccgcacccctggcgagcgtgctattccagtcgla 840
Db      3450 CTCTACTGCTTCAACCCCGTCTACCCGCACCCCTGGCGACGCTGCTATTCAGTCGTA 3509

QY      841 ttgggtgactagtcgggtgtctctgtgtgagcgtgtgtgacttgagccgttgcgctg 900
Db      3510 TTGGGTGACTAGTCGCGGCTCTTGGTGGCGGCTTGACTTGAGCCGTTGCCGCTG 3569

QY      901 cgcgtgttgagcatcttctgtggggtgagagacagcagcaacactccgagatcaatgtc 960
Db      3570 CCGCTGTGAGCATCTTTGTGGGCGGTGAAGACAGCAACAACCTCCGGATCAATGTC 3629

QY      961 tggctggcgtgtgtcgagatgattcaagatcgcccttgctgagcagcagcccgcaat 1020
Db      3630 TGCGTGGCGGTGCTGAGATGATTTCAAGATCGGCTTGCGTGGCATCGCCCGCAAT 3689

QY      1021 accgccttlaacctggttataccctatcaccaagcgcgcttlaagcggttagcgcc 1080
Db      3690 ACCGCCTTAAACCTGTGTTATCCCTCTATCAACAGCGCGCTTTACGGCGTTGAGCGCC 3749

QY      1081 tactcgtcccgctggaagtcgaggttagggcgagcactagtggttgagcgcttcgct 1140
Db      3750 TACTCCGTCCCGCTGGAAGTCGCGGTGAGGGCGGACTACTGGGCTTGACGGCTTCGCT 3809

QY      1141 tggctgtctgtgtcaacggtgagcggtgagcggtgagcaggttagccgactgagcgagat 1200
Db      3810 TGCGTCTGCTGCTGTCACGGCGGTGACGGCGGCTGCGGCAAGTGAGCCGACTGCGCGCAT 3869

QY      1201 cgcaatcccaagcctttgtgtgatgtgctagcttgccggtttgaggaatgtggt 1260
Db      3870 CGCAATCCCAAGCCTTTTGTGTGATGGCTTACCTTGCGCGGTTTGGCAGCAATGCTGGGT 3929

QY      1261 caaggtctgttgataccgtgtctatcgaccggaagcagtaagctcgtgtgtcgt 1320
Db      3930 CACGCTGTGTTGATACCGTCTCTATACGACCGGAAGCCAGTACGCTGTGGCTGTGT 3989

QY      1321 atggagcagatcgaggtttctgagcagcccaaccttccaagcaactccctcagaagcc 1380
Db      3990 ATTGAGCGATCGCGAGTTCCTGCGAGCCCAACCTTCCAAGCAACTCCCTCCAGAAGCC 4049

QY      1381 gagcattcagacgaaaaaatgtag 1404
Db      4050 GAGCATTCAGACGAAAAAATGTAG 4073
```

RESULT 2

D90911 LOCUS 143051 bp DNA BCT 04-JUL-2001

DEFINITION Synechocystis sp. PCC 6803 DNA, complete genome, section:13/27, 1576593-1719643.

ACCESSION D90911 AB001339 BA000022

VERSION D90911.1 GI:1653083

KEYWORDS

SOURCE Synechocystis sp. PCC 6803 (strain:PCC6803) DNA.

ORGANISM Synechocystis sp. PCC 6803

REFERENCE 1 (bases 1 to 143051)

AUTHORS kaneko,T., Tanaka,A., Sato,S., Kotani,H., Sazuka,T., Miyajima,N., Sugiyura,M. and Tabata,S.

TITLE Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb region from map positions 64% to 92% of the genome

JOURNAL DNA Res. 2 (4), 153-166 (1995)

MEDLINE 96127529

REFERENCE 2 (bases 1 to 143051)

AUTHORS kaneko,T., Sato,S., Kotani,H., Tanaka,A., Asamizu,E., Nakamura,Y., Miyajima,N., Hirose,M., Sugiyura,M., Sasamoto,S., Kimura,T., Hosouchi,T., Matsuno,A., Mureki,A., Nakazaki,N., Naruo,K.,

TITLE Okumura,S., Shimpo,S., Takeuchi,C., Wada,T., Watanabe,A., Yamada,M., Yasuda,M. and Tabata,S.
Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions
JOURNAL DNA Res. 3 (3), 109-136 (1996)
MEDLINE 97061201
REFERENCE 3 (bases 1 to 143051)
AUTHORS Tabata,S.
TITLE Direct Submission
JOURNAL Submitted (28-JUN-1996) Satoshi Tabata, Kazusa DNA Research Institute, The First Laboratory for Plant Gene Research, Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
(E-mail:tabata@kazusa.or.jp, URL:http://www.kazusa.or.jp/cyano/, Tel:81-438-52-3933(ex.2330), Fax:81-438-52-3934)
COMMENT Potential protein coding regions were assigned on the basis of similarity search of the ORFs and Genemark analysis.
FEATURES
source location/Qualifiers
1. .143051
/organism="Synechocystis sp. PCC 6803"
/strain="PCC6803"
/db_xref="taxon:1148"
/note="synonym:Synechocystis PCC6803"
173. .1960
/gene="slr2105"
173. .1960
/gene="slr2105"
/note="ORF_ID:slr2105
unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAA18000.1"
/db_xref="GI:1653084"
/translation="MARLTALFRTYGYAFLPAIAGLAGIALGVTGVMKLYLGLM AIAVIEFFLMTWLMQSGFLGRSTQSSANYASTLAVLAITLINYIAFRYPVAM DLTENQRFLLAPQOSLVALEQPLKWLIFDDASEKDRLLQNYQRFSDQSFSEFVH PDRRPKLVQDENVQNKGVFIIEYDKQQLVQNLSEQQREDLSEIQLTNAITAKAQOE IPQVYLVOGHGELPLNPPRGISTAVTALLENLGYVQPLNVTDGIPEDADAVIAG PTRKFLAPEIATLKQYTDQGNLLVMVSPREADAGLEDLLEPMGVSFDPRLVIDLSGAG CVEGLDPTVPFIENRYGNHPIQKLOGAIAIFLYRPVATKEOPHIQATTVIASDLMY ATQLISEDLPNPETDLGPDVAVALTRGEKEYSSAKEMETGNTEPEETNPSPSP VSPSPSTGNGTNPPEAMPEDNSQAGDRESPAGENEDSNGDAQTAQNGDPELPP AKMVVVGSAAFATNANFNNAVNGDIFLNSVQWLAENVQPLNVRAKDPTDRINLGV R R A T A V L G W L T W V I V P L I G L G I A V E T W M R Q R "
complement(2133. .2651)
/gene="apt"
complement(2133. .2651)
/gene="apt"
/note="ORF_ID:s111430"
/codon_start=1
/transl_table=11
/product="adenine phosphoribosyltransferase"
/protein_id="BAA18001.1"
/db_xref="GI:1653085"
/translation="MDLKALIRDIPDPKPGIMERDITLLNSPEGLRYTIDSLVEQC ESQELVPDHVVGMEGRGFLFGMPLAYQMNAGFIPVRKPGKLPAPVHRVEYDLEYGKDS LEIHQDAVAPHHRVLYDDLIATGTAKATAELLTKLGCEVLGFAFIIEALALNGROC LPDLPIISLVEY"
complement(3032. .3946)
/gene="s111429"
complement(3032. .3946)
/gene="s111429"
/note="ORF_ID:s111429
unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAA18002.1"
/db_xref="GI:1653086"
/translation="MILYLPMLIGAGATATFVWGPNKIKNAWDEEDKKKEEKRRKRE ARNTEAKAIWDAYERGEMLVFEKERISQEMNOYSKNILSEDRKNTKNLEEDIEHIEK YIEQSKTIYSVWEYVDSQSLIEDSHYDTRICKMEKENNEKRNMKKMESSNYINS HIFLYDTLDLKKECKEQOKLIEDGMSIDKMDGVDFEKVILTHFTFMGRGSLTOQT QDFGADLIIMKDRKKYVIQAKRSASAVSKAIOEILGAKGYDADIAYVTNNSYTEN

gene complement(3994. .4209)
/gene="ss12789"
complement(3994. .4209)
/gene="ss12789"
/note="ORF_ID:ss12789"
/codon_start=1
/transl_table=11
/product="resolvase"
/protein_id="BAA18003.1"
/db_xref="GI:1653087"
/translation="MGQFEREMMLERQREGVAKAEKKYEARRPISDGVPEQVTVLRH GRSKAYVARELGISQASVYRIINLCSA"
complement(4340. .5218)
/gene="s111428"
complement(4340. .5218)
/gene="s111428"
/note="ORF_ID:s111428"
/codon_start=1
/transl_table=11
/product="p3 protein"
/protein_id="BAA18004.1"
/db_xref="GI:1653088"
/translation="MESNFLTTFILPLALFLIMFGMGLGLTIKDENRIWLEPKAVAIG LIAQLVLEPMWFGLSLSLSPELAVGLMILAACPGGSVSNVITYLLKGNVALSITL TAISLSVITITIPLVNLANYFMGQFALQPLFKTVLQAVITIIIPVSLGMIFYHF LPKIAMVMEKNVKWLSLFLGIIIVAILVKERENLADFFLOVGVTNLNLTMGIGY GIATLACGLAERKSTITEVGIQNGTLAIAIATTLINAPITMAIPAIIYSVMFLTSGI FAGLTKAKIRFGLKST"
5614. .6606
/gene="slr1503"
5614. .6606
/gene="slr1503"
/note="ORF_ID:slr1503
unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAA18005.1"
/db_xref="GI:1653089"
/translation="MTISTDTSLTIIMPAMGCGTWAMGNKLLMGYQAPAMNGELRVER ECVAAGITFFDGTDSYGTGRLOGSEKLLGQFSQAYEGLNODKICLATKLAPYPRILT PQAMVGAQMASQORLQRPIDLVQHMSTANYAPWQEQPLKGLVQLCQRGHAKAIGLS NFGPKRLQKAYDFCQNEGVSIVSLQVYSLSTDPVSKGLKEQLGKILAVSPL TLGLTLGRKYGPQGPFPFGIRGLFRLLPKIQPLDITQAIAREREKTMAGVALNMCI CGGTMPIPGAKNCQOVODNLGALGQWQLDRGEMDELNNQAAALPQTMVONIQSR"
complement(6657. .7907)
/gene="hob"
complement(6657. .7907)
/gene="hob"
/note="ORF_ID:s111427
gene:hob or htrH"
/codon_start=1
/transl_table=11
/product="protease; hob"
/protein_id="BAA18006.1"
/db_xref="GI:1653090"
/translation="MAIHLKASHLCVAVLLLLFGAIGAACGYYLLSSGQNHSSPDSF VNTSPQSLTPAPVESNYRSALPTLPRSAQDDQELNFIARVQKIGPAVVRIDSERTA VSQCGPMGDQPFRRFEGEEMPNDPREQGTSGFILLSGDEVLTNAHVEGASTVK VTLKDGSLVEGKVMGIDTMTDAVAVKVEAENLPVEIGQSDRLQGEWAIAIGNPLGL DNTVTVGIIISALGRSSSEGVDPKRYRFIQTDAAINPGNSGGPILNAKGEVIGVNTAI RADAGGLGFAIPLOTQONVAENLEFTKGMEHPYLGIMHVTLPPEMTKOLRTSGELPAG VTADTGVLIIOVSPSPAQAGLAPGDIILEVGMGVKATVDQERVEVSQIGEPILAI AVKRGQKQOMAAVVRPFPEDIGQ"
complement(8021. .8491)
/gene="s111426"
complement(8021. .8491)
/gene="s111426"
/note="ORF_ID:s111426
unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAA18007.1"

```

/db_xref="GI:1653091"
/translation="MMKSLRGIALSVSLMATGGYGLTSAQVAPGDDIPTVQALD
TPVDVDTIDQDOLITWICKNGSSVAVEVKELGWOETWDOQWACEONIPSYDPS
DATFSCETNATMGLITVFWLEGEKGAKEMTSWMNNLANQNVCTHGQSNPWE"
complement(8617..10470)
/gene="pros"
complement(8617..10470)
/gene="pros"
/feature="ORF_ID:s111425
gene:pros or drpa"
/codon_start=1
/transl_table=11
/product="proline-tRNA ligase"
/protein_id="BA18008.1"
/db_xref="GI:1653092"
/translation="MWDNSVRHLPRLLMRLSOMLFVTLRBDPADAEIPSHKLVRAG
FIRLGSGLAYLPLIWRVLEKVKRIVQEEEMNRGAOECLLPOLPSELWKMGRMDT
YTESGIMFALRDLERELGPTHEEVITAIKEMIRSRQLPVLNXYQTKERDEI
RPRGLMRGREFIMKDAYSPHDEASLKATYGAMDQAYRNIFSRGLDFRPVDADSGA
IGSGSQEEMVLADAGEDEILYTEDGLYSANVEKAVSLADAKPSFSGYKKOSAPNT
ATIAKQCQLCDASNVKKNVLAQVYNSGKAVLVLSIRGQEVNEVKLTNELKLA
PHYESTNLSLTVPDEKQKNAKPLPLGISPSIDDIATINKVADKFLRLVDT
AVDLNFAATGADAVDYHVTGAWSTDFTLPLLVNDRKALAGDRPHDPTGELKTARG
IEVGHIFQGLTKYSQAMGATFTNEQGEKPLVMGCYIGVSRLAQAAVEOSYDKGIL
WPAIAPYEYVIVPNVGDSEDMQVAEQIYTEFOAAEVYVLLDDRNERAGVKFKDSEL
IGIPYRVVTGKSLKQGLLEVYKRAKDSVONLAIAAVVPYTLAWIKAEEKTS"
10639..11235
/gene="slr1505"
10639..11235
/gene="slr1505"
/feature="ORF_ID:s1r1505
unknown protein"
/codon_start=1

```

```

Query Match      23.2%; Score 326.4; DB 1; Length 143051;
Best Local Similarity 52.7%; Pred. No. 6.2e-66;
Matches 708; Conservative 0; Mismatches 636; Indels 0; Gaps 0;

QY      1 atgactgtctgcaaacctctgactttgcccattaccaccccaacagtgggccacagc 60
Db      28283 ATCTTAATCTGGCGATCGCTGATGTTGGCGGTTTTCCTCCCAAGGAATGGCGCGGC 28342

QY      61 agttctctgcatcgctgtttgagcagctgagcttgcgcgccaccagcagctgttg 120
Db      28343 AGTGTGCTCCATCGTTTGGTGGGCTGGGAGACAGATTGGATACAGCCTAGTGTCTCTG 28402

QY      121 gttgtgtctgaggcactggtgtgctcttctgtctgtctgtctacagtttcgctcgtt 180
Db      28403 CCCCACTTCGAGGCATTTGGGTACGGCTCACTGCGCAATATTTTATTCGGCTCCCTTC 28462

QY      181 gtgccagttccgcccctagagttggtgagcagcgatcgcgccctattggtccctgtc 240
Db      28463 ACCTCCACCAACCATGTTGGGCATTTTATGCTGCTCTGTGAGACCTTTTGGGCTCTGCTG 28522

QY      241 tcgctgacagatatcgatctcgcgcaagcaacccccattcactgctgctgctgtctac 300
Db      28523 ACCTTTGTCTGATCAACCAAGGAGGCTTTGACTCCCATCCATCATGTTTATGTTTGGCTTAC 28582

QY      301 tgggagctgcatgcccctagcaacggtgactcaccgcgtacgctgagctttagttggg 360
Db      28583 TGGTGCATTTTCGGCGATCGCCGTGGGATTTTCTCCGGTAAATAATGCGCGCGCGCTCGGGG 28642

QY      361 ctgaccacactgagcctctacgtgtgttlltgcctagcggtcgggttctccgcaat 420
Db      28643 TTAGCGAAATTAACAGCTAATTTATGCTGTTTCTACTGCGCGGAGGTTATTGCAAAAC 28702

QY      421 cccgctcgcgatcgctgtctctcggtcgctgctgatacatcgcttlltgcagtgtc 480
Db      28703 AAACAATGTTGAACCGGTTAGTAACGTTGTTTACTGTAGGGCTATTTGTTGGGAGT 28762

QY      481 tacggctcaaccaatgatctacggcgttgaagagctggtgagcttgggtgatacgcaac 540
Db      28763 TACGGTCTGCGACACAGGTGACGGGGTAGAACAAGTTAGCCACTTGGAAATGACCCACC 28822

```

```

QY      541 tcggttcgcgaacttcacctcaacggtttacagctatctggtcaaccccaacctgtgct 600
Db      28823 TCTACCTTGCCCAAGGCCACTAGAGGTATATAGCTTTTATAGTAATCCCAATCTCTTGGCG 28882

QY      601 gctatctgtagccagcagcagctcttcttcagcagcagcagcagcagcagcagcagc 660
Db      28883 GCTTACCTGTGCCCCATGACGAGGTTTGAGCTTGAGTGCCCTGTGTGTATGGCAGCGTGG 28942

QY      661 ctccccaagctgctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 720
Db      28943 TGGCCCAACTGCTGGGAGACCAACCATGATGTTTAACTACTCTGCTCTTTTAACT 29002

QY      721 tacagtcgcggtgtgctgtgctgtgttctgcacatgattlltctctggtggttataagg 780
Db      29003 CAGAGCCCGGGCGGTTGGCTAGCAGTGTGCTGCCCTGGAGCTTACCTTCTGCGCCCTTGT 29062

QY      781 ctctactgtttcaaccccgctcaccgcagcagcagcagcagcagcagcagcagcagcagc 840
Db      29063 TACTTCTGTGTGTTACCCCAATTACCCAAATTTTGGCAACGCTGTCTTTGCCCTGTGCG 29122

QY      841 ttggtgtgacagtcgagcgtgtctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 900
Db      29123 ATCGCCGTGGCGGCTTATATATAGGTGGGCGACGCTTGATGTGGGTGGAAACGATTCGACTC 29182

QY      901 cgcgtgtcagcatcttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 960
Db      29183 AGGGCCATGACCATTTTGTGCTGGCGGGAAGACAGACAGCATATATTTCCGCATCAATGTT 29242

QY      961 tggctgtgagtgctgtcagatgatcagaatcagcagcagcagcagcagcagcagcagc 1020
Db      29243 TGGGAAGGGGTAATAAGCATGATCGGAGCCGCCCTATCATTTGGCAATTGGCCAGGTAAC 29302

QY      1021 accgccttaacctggttataccctctatacaacagcgagccttaacgagcttgagcgc 1080
Db      29303 GAAGCCTTTAAACCAATTTATCTTACTATATGCGGCGCCCTTCAACGCCCTGAGTGCC 29362

QY      1081 tactcgtcccgctgtgaagtcggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1140
Db      29363 TATTCATTTTACCTAGAAATTTGCTGGAACAGCGGTGTGTGTGTGTGTGTGTGTGTGTGT 29422

QY      1141 tggctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1200
Db      29423 TGGCTGTGGCGCTTACCTAGGCAAGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 29482

QY      1201 cgcacccccaagcctlltgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1260
Db      29483 CTCGCCCGGAAGGCATCTGATATATGGGGGCTTTAGCGCGCATCATCGTGTGTGTGTGTGT 29542

QY      1261 caggtctgtttgtataccgtgtctctatcgaccggaagcagtaagcgtctgtgtgtgtgt 1320
Db      29543 CACGGCATGTGATATACAGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 29602

QY      1321 attgagcagatcgaggtttcttg 1344
Db      29603 GTGGCCATTTGTTGCTAGTCACTGG 29626

```

```

RESULT      3
LOCUS      166494      7218 bp      DNA      PAT      28-DEC-1997
DEFINITION      Sequence 14 from patent US 5670367.
ACCESSION      166494
VERSION      166494.1      GI:2724471
KEYWORDS
SOURCE      Unknown.
ORGANISM      Unclassified.
REFERENCE      1 (bases 1 to 7218)
AUTHORS      Dörner, F., Schefflinger, F. and Falkner, F. Gunter.
TITLE      Recombinant fowlpox virus
JOURNAL      Patent: US 5670367-A 14 23-SEP-1997;
FEATURES      Location/Qualifiers

```



```
/note="isoniazid and ethionamide target protein"
/codon_start=1
/transl_table=11
/product="inhA peptide (AA 1-269)"
/protein_id="AAC43210.1"
/db_xref="GI:407314"
/translation="MTGLDGRILVSGIIPDSSIAFHIAVRAOEOGAOLVLTGEDRL
RLIQRITDRLPAPAKAPLLELDVQNEHLASLAGRVTEAIGAGNKLGVVHSGFMPQTG
MGINPFADAPYADVSKGIHISAYSASMAKALLPIMNPGSIVGMEFDPSPAMPAYNW
MTVAKSALSVNRFVAREAGKYGVRSNLVAAGPIRTLAMSALVIGALGEEGAQIQLL
EEGWDQRAPIGMNMKDATPVAKTVCALLSDMLPATTTGDIYADGAHTQLL"
BASE COUNT      153 a      256 c      285 g      138 t
ORIGIN

Query Match      3.4%; Score 47.4; DB 1; Length 832;
Best Local Similarity 54.2%; Pred. No. 1;
Matches 96; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY  1070 cgttgagcgctactccgctccgcctggaagtcgcggttgagcggaactagtggcttga 1129
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   77  CGTCGATCGCGTTTCACATCGCACGGGTAGCCACGAGCAGGCGCCCAAGCTGTGCTCA 136

QY  1130 cggccttcgcttgctgtgctgtgcaagcgcggtgcaagcgcggaagtgaagccgac 1189
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   137 CCGGGTTGACCGGGCTGCGGCTGATTCAGCGCATCACCGACCGGCTGCCGCAAGGCC 196

QY  1190 tgcgcgcgcatcgcaatccccaagccttggtagtgtagcttgccggttgg 1246
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   197 CGCTGCTCGAAGTCGACGTCGCAAAAGAGAGACACCTGGCCAGCTTGCCGCGCGG 253

RESULT 6
LOCUS      MBU41388      1789 bp      DNA      BCT      16-JAN-1996
DEFINITION Mycobacterium bovis putative ketoacyl ACP reductase and enoyl ACP
reductase (inhA) gene, complete cds.
ACCESSION      U41388
VERSION      U41388.1 GI:1155268
KEYWORDS
SOURCE
ORGANISM      Mycobacterium bovis strain=Wag201.
Mycobacterium bovis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE
AUTHORS      Banerjee,A., Dubnau,E., Quemard,A., Balasubramanian,V., Um,K.S.,
WILSON,T., COLLINS,D., de Lisle,G. and Jacobs,W.R. Jr.
inhA, a gene encoding a target for isoniazid and ethionamide in
Mycobacterium tuberculosis
Science 263 (5144), 227-230 (1994)
JOURNAL
MEDLINE      94112548
REFERENCE
AUTHORS      Wilson,T.M., de Lisle,G.W. and Collins,D.M.
Efect of inhA and katG on isoniazid resistance and virulence of
Mycobacterium bovis
Mol. Microbiol. 15 (6), 1009-1015 (1995)
JOURNAL
MEDLINE      95349387
REFERENCE
AUTHORS      Collins,D.M.
Direct Submission
Submitted (27-NOV-1995) D.M. Collins, Animal Health Division,
Agresearch, Ward Street, Upper Hutt, P.O. Box 40063, New Zealand
FEATURES
source
1..1789
/organism="Mycobacterium bovis"
/strain="Wag201"
/db_xref="taxon:1765"
1..162
/gene="inhA"
1..162
/promoter
1..162
/gene="inhA"
163..906
CDS
```

```
/note="putative ketoacyl ACP reductase; Method:
conceptual translation supplied by author."
/codon_start=1
/transl_table=11
/product="enoyl ACP reductase"
/protein_id="AAB60182.1"
/db_xref="GI:1155269"
/translation="MTATATEGAKPPEVRSVLVTGNRGIGLAIAQRLAADGHRVAV
THRSGAPKGLFGECDVTDSADVAFTVAEEHOGPVEVLVSNAGLSADAFLMRTE
EKFEKVINMLTGAFRVAQRASQMRNKGFMIFIGSVSGSMGIGNQANYAASKAGV
IGMARSIARELSKANVTANVAPGYITDITRALDERIQGALQFIPAKRVGTPAEVA
GVVSFLASEDASYISGAVIPVDGGMNGH"
925..1734
/codon_start=1
/transl_table=11
/function="mycolic acid synthesis"
/product="enoyl ACP reductase"
/protein_id="AAB60183.1"
/db_xref="GI:1155270"
/translation="MTGLDGRILVSGIITDSSIAFHIAVRAOEOGAOLVLTGEDRL
RLIQRITDRLPAPAKAPLLELDVQNEHLASLAGRVTEAIGAGNKLGVVHSGFMPQTG
MGINPFADAPYADVSKGIHISAYSASMAKALLPIMNPGSIVGMEFDPSPAMPAYNW
MTVAKSALSVNRFVAREAGKYGVRSNLVAAGPIRTLAMSALVIGALGEEGAQIQLL
EEGWDQRAPIGMNMKDATPVAKTVCALLSDMLPATTTGDIYADGAHTQLL"
1204
variation
/note="causes a change in the encoded protein from serine
to alanine. Expression of this altered protein in
Mycobacterium bovis and Mycobacterium smegmatis confers
isoniazid resistance."
/replace="G"
BASE COUNT      340 a      550 c      602 g      297 t
ORIGIN

Query Match      3.4%; Score 47.4; DB 1; Length 1789;
Best Local Similarity 54.2%; Pred. No. 0.93;
Matches 96; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY  1070 cgttgagcgctactccgctccgcctggaagtcgcggttgagcggaactagtggcttga 1129
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   980 CGTCGATCGCGTTTCACATCGCACGGGTAGCCACGAGCAGGCGCCCAAGCTGTGCTCA 1039

QY  1130 cggccttcgcttgctgtgctgtgcaagcgcggtgcaagcgcggaagtgaagccgac 1189
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   1040 CCGGGTTGACCGGGCTGCGGCTGATTCAGCGCATCACCGACCGGCTGCCGCAAGGCC 1099

QY  1190 tgcgcgcgcatcgcaatccccaagccttggtagtgtagcttgccggttgg 1246
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   1100 CGCTGCTCGAAGTCGACGTCGCAAAAGAGAGACACCTGGCCAGCTTGCCGCGCGG 1156

RESULT 7
LOCUS      I73546      2232 bp      DNA      PAT      03-APR-1998
DEFINITION      Sequence 12 from patent US 5686590.
ACCESSION      I73546
VERSION      I73546.1 GI:3009687
KEYWORDS
SOURCE
ORGANISM      Unknown.
Unclassified.
REFERENCE
AUTHORS      1 (bases 1 to 2232)
Jacobs,W.R. Jr., Collins,D.Michael, Banerjee,A., de Lisle,G.William
and Wilson,T.Mary.
Methods and compositions for detecting and treating mycobacterial
infections using an INHA gene
Patent: US 5686590-A 12 11-NOV-1997;
JOURNAL
MEDLINE      Location/Qualifiers
1..2232
/organism="unknown"
1..2232
/promoter
411 a      696 c      751 g      374 t
BASE COUNT
ORIGIN
```


KEYWORDS	unknown.
SOURCE	unknown.
ORGANISM	unclassified.
REFERENCE	1 (bases 1 to 3120)
AUTHORS	Sacchetti, J., Blanchard, J. and Jacobs, W.R. Jr.
TITLE	Method and compounds for inhibiting lipid biosynthesis of bacteria and plants
JOURNAL	Patent: US 5702935-A 1 30-DEC-1997;
FEATURES	Location/Qualifiers 1..3120
BASE COUNT	566 a 995 c 1037 g 522 t
ORIGIN	/organism="unknown"
Query Match	
Best Local Similarity	3.4%; Score 47.4; DB 6; Length 3120;
Matches	96; Conservativity 0; Mismatches 81; Indels 0; Gaps 0
QY	1070 cgttagcgcactactccgtccgcgtggaagtcgcggttgaggcgagactagtgtga 1129
DB	1041 CGTCGATCGCGTTTCACATCGCACGGGTAGCCACGAGCAGGGCGCCAGCTGTGCTCA 1100
QY	1130 cgccttcgcttgctgtgtgtgtgacgagcggtgacgagcggtgcgcagtgtagcgac 1189
DB	1101 CCGGCTTCGACCGCGCTGCGGCTGATTCAGCGCATCACCGACCGGCTGCCGCAAGGCC 1160
QY	1190 tgcgcgcgcagtcgcacatccccaagcctttgttgatgtgtagctgtgacggttg 1246
DB	1161 CGCTGCTCGAAGTTCGACGTGCACAAACGAGGACACCTGGCCAGCTGGCCGGCGG 1217
RESULT	12
LOCUS	AE007022 20113 bp DNA BCT 27-APR-2001
DEFINITION	Mycobacterium tuberculosis CDC1551, section 108 of 280 of the complete genome.
ACCESSION	AE007022 AE000516
VERSION	AE007022.1 GI:13881145
KEYWORDS	
SOURCE	Mycobacterium tuberculosis CDC1551.
ORGANISM	Mycobacterium tuberculosis CDC1551
REFERENCE	Bacteria: Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.
AUTHORS	1 (bases 1 to 20113)
REFERENCE	Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouiri, H., Gill, J., Mikula, A. and Bishai, W.
TITLE	Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 20113)
AUTHORS	Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouiri, H., Gill, J., Mikula, A. and Bishai, W.
TITLE	Direct Submission
JOURNAL	Submitted (25-APR-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
FEATURES	Location/Qualifiers 1..20113
source	/organism="Mycobacterium tuberculosis CDC1551" /strain="CDC1551" /db_xref="taxon:83331" /note="clinical strain" 100..1233 /gene="MT1526" 100..1233
CDS	

/gene="MT1526"
 /note="similar to GP:5458705; identified by sequence
 similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="magnesium chelataase, putative"
 /protein_id="AAK45791.1"
 /db_xref="GI:13881146"
 /translation="MTSAGGFFAGAGCYQTPGCHSASPANEAPPGAGEGLAEVHTLE
 RAIFEVKRIIVGODOLVERMIVGLSKGHVLLGEGVAKTAVETFARYGGTESRI
 QFTPLDLPPTDIIGTRIRYRQGEEDTELGPVVAANELADEINRAPAKVQSALLEVMOE
 RHVSIIGRTEPMPSPFLMATQNPPIEHGVPVLPFAORDRFLFKINVGYSPEEREI
 IYRMGVTPPQAKOILSTGDLRLQEIANNFVHALVDYIVRVFATRKPEQLGMNDV
 KSVAVFGASPRASLGIIAARSALIVRGDVPIDQYIEVLPVLRHRLVLTLDALD
 EISPELVINRVLQTVLPPQVNAVPPQGHSPVPMQAAAAASGR"
 1272. .2183
 /gene="MT1527"
 1272. .2183
 /gene="MT1527"
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAK45792.1"
 /db_xref="GI:13881147"
 /translation="MLRGDIDDPKLAALRTLELVKQKLDGVLHGDLGLIPGSGE
 PGESRLYQPGDDVRRMDWAVTARTTPHVRQMLADREFTLVLDVMSASIDFGTACCE
 KRDLAVAAAAAITEFLNSGGGRGALIIANGAAMTRVPARTGRQHTMLRTIATMPQ
 PAGVRGDLAVAI DALRRPERRGMAV IISDFLGINMMRPLRAIARHEVLAIEVLDP
 RDVELPDVGDVVLQDAESGVVREESIDPALRDPARA AAAHRADVARTIRGCGAPLLS
 LRTDRMDADIVRFVASRRGALAGHQ"
 2194. .3201
 /gene="MT1528"
 2194. .3201
 /gene="MT1528"
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAK45793.1"
 /db_xref="GI:13881148"
 /translation="MTPLPLGPMYTLGFAHSWFFLFLVAVAGLVALYILMQLARQRM
 LRFANMELLESVAPKRPKRSHVPAIILVLSLFTIAMAGPTHVRIIPNRVAVMLV
 IDVSQSMRATDVEPSRMVAAQEA AKQFADELTPGINLGLIAYAGTAVLVSPPTNREA
 TKNALDKLQFADRTATGCAITFALQAIATVCAVIGGGDXPPARIVLESDKETMPTN
 PDNPGAVYTAARTAKDQGVPISTISFGTPYGEVEINDORQVPVDDETMKKVAQLSGG
 NSYNAATLAEIRAVYSSLDQQIQGYETIKGDAVGMRLGALALALALALLINRRLP
 T"
 complement(3274. .4152)
 /gene="MT1529"
 complement(3274. .4152)
 /gene="MT1529"
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAK45794.1"
 /db_xref="GI:13881149"
 /translation="MTFVGVQTQHTSTMTDPFLGSEALAGVLTPTYELRSRYVALHKDV
 YVPGVELTQALRAKALWRSRRGVLAGYSASFHGA KAWIDADIPAIIDTNRRAP
 GLQVVEERIEPDEICVTEGMRYVTPERTALDLTSRFLPDPAVAADVALIQATDLKVA
 VEPLEIRYRGRGKMKARAALADLVDGAAQSPKETWLRLLIRAGCFPRQTOIAVRNEW
 GMAEFAHLDGMQDIKVAAEYDGDHLLTSRYHYRKDILRHEKVQHRGMIIVRVAAEDH
 PADIIIRVGEARAFRA"
 4257. .5000
 /gene="MT1530"
 4257. .5000
 /gene="MT1530"
 /note="similar to PID:1155269; identified by sequence
 similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="3-oxoacyl-(acyl-carrier-protein) reductase"

```
/protein_id="AAK45795.1"
/db_xref="GI:13881150"
/translation="MTATATEGAKPPFVSRSLVTGNGRIGLAIARLADGKHVAV
THRSGAPKGLFGVECDVTDSDADVDRAFVAVEHQGVFVVLVSNAGLSADAFLMRME
EKEKVINANLTGAFRYAQRASRSMQRNKGFMIFISVSGSWGIGNQANYAASKAGV
IGMARSIARELSKANVTANVAVPGYIDTDMTRALDERIQGALQIFPAKRVGTPAEVA
GVVSFLASEDASYISGAVIPVDGGMGMGH"
5019. .5828
/gene="MT1531"
5019. .5828
/gene="MT1531"
/notes="similar to GB:U02530 SP:P42829 PID:437204;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="enoyl-(acyl-carrier-protein) reductase"
/protein_id="AAK45796.1"
/db_xref="GI:13881151"
/translation="MTGLLDGKRILVSGIITDSSIAFHIAFYAOEQALVLTGFDRL
RLIQRITDRLPAKAPLLELDVQNEEHLASLAGRYTEAIGANKLDGVHSGEMPQTG
MGINPFEDAPYADVSKGIHISAYSASMAKALPIMNPGSIVGMDPSPAMPAYNW
MTVAKSALESVNRFAVREAGKGVSRNLVAGPIRTLMSAIVGALGEEAGAQIQLL
EEGWDQRAPIGWNMKDTPVAKTVCALLSDWLPAFTGDIYADGAHTQLL"
5834. .6868
/gene="MT1532"
5834. .6868
/gene="MT1532"
/notes="similar to PID:2347006; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="ferriochelatase"
/protein_id="AAK45797.1"
/db_xref="GI:13881152"
/translation="MQFDVALLLSFGCPGPEQVRPFLNVTGRGVPAERLDAVAEH
YLHFGVSPINGINRILIALEAQOELPVYFGNRMPEPYVEDAVTAMRDNGVRAAVF
ATSAMSGYSSCTQYVEDIARAKRAAGRDAPELVKLRPYFDHPLFVEMFADATTAAT
VRGDARLVFTAHSIPTADRCGPPLYSRQVAVATRLVAAAAGYCDFLMAQSRSGP
QVFWLEPDVTDLTGLAGAGINAVIVCPITGFVADHIEVWDLDELRLQAEAGIAYA
RASTPNADPRFARLARGHIDELRYGRIPARVSGPDVPGLSSINGQCPRPCHVASV
SPARPSAGSP"
complement(6834. .7700)
/gene="MT1533"
complement(6834. .7700)
/gene="MT1533"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK45798.1"
/db_xref="GI:13881153"
/translation="MNCPSVSLSIWANAWLAGKAAPDDVLDALSLMWPQSVAAAYDAV
AAGHTGLPMPDVHDAGTVSLIQLTRAAGVRRRLRGITINVLPVPGDVRGLAGTOFEH
DALAGEAVIVANPEPDSAVGLVPEFSYGDVDEAAQSEPLPELICALSMNYSILGA
PVLEHYELGDAEYALRAVRSAAEALSTIGLSSDVANPRGLVEQLLESSRQHRVPDH
APSRALRVLENAHVDAILIIVSAGLSRLPIGTOSLSDAQRATDALRPLTAVVRSARMS
AVTALHSAWPD"
7758. .8192
/gene="MT1533.1"
7758. .8192
/gene="MT1533.1"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK45799.1"
/db_xref="GI:13881154"
/translation="MPVALTWLIAALVLVGAELTGMFLMLGGGALAASVSSWLLA
WPMWADGAVFLIVSVLLIVLRPAVRRLTOTKGVLQIEALEGKKAIVLGRVARDGG
QVKLDGQVWTARPLNDGVFEPPGDSVTVVOIDGATAVVEKDY"
8214. .9359
/gene="MT1533.2"
8214. .9359
CDS
```

```
/gene="MT1533.2"
/notes="identified by match to protein family HMM"
/codon_start=1
/transl_table=11
/product="SPFH domain/Band 7 family protein"
/protein_id="AAK45800.1"
/db_xref="GI:13881155"
/translation="MOGAVAGLVFLAVLVIFAIIVAKSVALIPQAEAVIERLGRYS
RTVSGQLTLVLPFIDRVARAVDLREYVSPPOPVITTEDNLITDVIYFQVTVPOA
AVEISNIYIVGEOLTTTLRLNVYGMITLEQTLTSRDQINOLRGVLDPEATGRWGLRV
ARVELRSIDPPPSIQASMEEKOMKADREKRAMILTAEGTREAIKQAEQKQAOILAAB
GAKQAAIILAEADROSRLRAQGERAAAYLQAOQAQAKAITEKTFPAIKAGRPTPEMLAY
QYLOTLEPMARGDANKVMVWVPSDFNALQGFTRLIGKPGEDGVFEPEPSVEDQPKHA
ADGDDAEYAGWFSSTDTPSIRAAVATAEAIARKPVEGSLGTPPRLTQ"
9369. .9725
/gene="MT1534"

gene

Query Match          3.4%; Score 47.4; DB 1; Length 20113;
Best Local Similarity 54.2%; Pred. No. 0.63;
Matches 96; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 1070 cgttgagcgccactccgtcccgctggaagtcgcggttgaggcgcnactactggcttga 1129
      ||| || ||| | | || | | || | | | || || || | | |||
Db 5074 CGTCGATCGCGCTTCACATCGCACGCGGTAGCCAGAGCAGAGCGCCGCGAGCTGCTCA 5133

QY 1130 cggccttcgcttgctgctgtgltcacggcggtgacggcggtgacgagtgagccgac 1189
      || || ||| | || || | || || | | || | | || | | ||
Db 5134 CCGGGTTGCGACCGCGCTGCGCTGATTCAGCGCATGCACCGACCGGCTGCCGGAAGGCC 5193

QY 1190 tgcgcgcgcatcgcaatccccaagccttgggtgatgtgctagcttgccggtttgg 1246
      || || || | | | | | || | | || | || || || || || ||
Db 5194 CGCTGCTCGAACTCGACGTGCAAAACGAGGAGGACACTGGCCAGCTTGCGCCGCCGGG 5250

RESULT 13
MTCY277
LOCUS MTCY277 38300 bp DNA BCT 03-AUG-2001
DEFINITION Mycobacterium tuberculosis H37Rv complete genome; segment 65/162.
ACCESSION Z79701 AL123456
VERSION 279701.1 GI:3261635
KEYWORDS
SOURCE Mycobacterium tuberculosis H37Rv.
ORGANISM Mycobacterium tuberculosis H37Rv
          Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
          Actinomycetales; Corynebacterineae; Mycobacteriaceae;
          Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE
  1 (bases 1 to 38300)
  Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,
  Harris,D., Gordon,S.V., Eiglmeier,K., Gas,S., Barry III,C.E.,
  Tekala,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T.,
  Connor,R., Davies,R., Devlin,K., Fellwell,T., Gentles,S.,
  Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Krogh,A., McLean,J.,
  Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A.,
  Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S.,
  Squares,S., Squires,R., Sulston,D.E., Taylor,K., Whitehead,S. and
  Barrell,B.G.
  Deciphering the biology of Mycobacterium tuberculosis from the
  complete genome sequence
  Nature. 393 (6685), 537-544 (1998)
JOURNAL Nature. 393 (6685), 537-544 (1998)
MEDLINE 98295987
REFERENCE 2 (bases 1 to 38300)
AUTHORS Parkhill,J.
TITLE Direct Submission
JOURNAL Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
On Jun 27, 1998 this sequence version replaced gi:1524225.
COMMENT
Notes:
Details of M. tuberculosis sequencing at the Sanger Centre are
available on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have
```

been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers. Gene prediction was based on a Hidden Markov Model of TB genes implemented in TBparse (Krogh) supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C.

CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

FEATURES

```
source
    location/Qualifiers
    1..38300
    /organism="Mycobacterium tuberculosis H37Rv"
    /strain="H37Rv"
    /db_xref="taxon:83332"
    <1..37608
    /organism="Mycobacterium tuberculosis H37Rv"
    /strain="H37Rv"
    /db_xref="taxon:83332"
    /clone="Y277"
    /gene="Rv1482c"
    complement(55..1074)
    /gene="Rv1482c"
    complement(55..1074)
    /gene="Rv1482c"
    /note="Rv1482c, (MTCY277.03c), len: 339. Highly similar to M. avium protein TR:O07396 (EMBL:AF002133) MAV346. (346 aa), fasta scores; E(): 0, 65.2% identity in 342 aa overlap, slight similarity to GRPE_ECOLI P09372 heat shock protein grpe (heat shock p (197 aa), fasta scores, opt: 139, E(): 0.012, (28.3% identity in 159 aa overlap). Also similar to Rv3517(MYV023.24), 59.0% identity in 273 aa overlap."
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein Rv1482c"
    /protein_id="CAB02032.1"
    /db_xref="GI:1524228"
    /db_xref="SPTREMBL:P71763"
    /translation="MGVWPLQWLWQSPROPVSPRGNNRYVHERNVAALISGNERFGVTLVTFVGVTQHTSTMTDPLGSEALAGVLTPELRSRYVALHKDVYVPQGVETLTAQLRKALMLRSRRGVLAGYASAFHGAKWIDADLPALITDNRRAAGLOVWEERIEPDEICVIEGMRVTPPERTALDLSRFPDPAVAADALIOATDLKVADEPLIERYRGRGMKKAARALDLVDGGAQSPKRTWLRLLLRAGFPRTQIOAVRNEMGMAEHLDMGMQDIKVAAEYDGDHLLTSRYHYRKDILRHEKVQHRYGWTWVVAEDHPADILRRVGEARAFRA"
    1024..1029
    /note="putative RBS, GATAGG, for Rv1483"
    1038..1781
    /gene="fabG1"
    1038..1781
    /gene="fabG1"
    /note="Rv1483, (MTCY277.04), fabG1, len: 247. fabG1, Probable 3-oxoacyl-[acyl-carrier protein] reductase, identical to M. bovis G1155269 putative ketoacyl acp reductase (247 aa), strong similarity to FABG_ECOLI P25716 3-oxoacyl-[acyl-carrier protein] reductase (244 aa), fasta scores, opt: 664, E(): 6.8e-35, (44.4% identity in 241 aa overlap; contains P500061short-chainalcohol dehydrogenase family signature"
    /codon_start=1
    /transl_table=11
    /product="fabG1"
    /protein_id="CAB02033.1"
    /db_xref="GI:1524229"
    /db_xref="SWISS-PROT:Q48930"
    /translation="MTATATEGAKPPFVSRSVLVTCNRCIGLAIARLAADGKRVAVTHRSGAPKGLFGVECDVTDSDAVDRAFTAVEEHQGPVEVLVSNAGLSADAFLEMTTEKEKVINANLTGAFRAVQARASRSMQNRNKGFMIFIGSVSGSWGICGNQANYASKAGVIGMARSIARELSKANVTANVAVPGYIDTDMTRALDERIQGALQIFPAKRVGTPAEVAGVSEFLASEDASYISGAVIPVDGGMGMGH"
    1455..1541
misc_feature
```

```
/gene="fabG1"
/note="P500061 Short-chain alcohol dehydrogenase family signature"
1790..1794
/note="potential RBS, AAGGA, for Rv1484"
1800..2609
/gene="inbA"
1800..2609
/gene="inbA"
/note="inbA"
/note="Rv1484, (MTCY277.05), inbA, len: 269. Identical to INHA_MYCTV P46533 enoyl-[acyl-carrier-protein] reductase and to M. bovis G1155270 enoyl acp reductase"
    /codon_start=1
    /transl_table=11
    /product="inbA"
    /protein_id="CAB02034.1"
    /db_xref="GI:1524230"
    /db_xref="SWISS-PROT:P46533"
    /translation="MTGLDGKRILVSGIIPDSIAFHIAVAAEQGAQLVLTGFDRLLRIQRTDRLPAKAPLLELDVONEEHLASLAGRYTEAIGAKNKLGVVHSIGFMPQTMGINPFEDAPYADVSKGTHISAYSASMAKALIPIMNPGSIVGMDFDSRAMPAYNMVAVKSALESVNRFEVAREAGKYGVRSNLVAAGPIRTLAMSIVGALGEAGAQIQLLEGDWDRAPICGMNMKDATPVAKTVCALLSDMLPATGTDIIVADGAHTQLL"
    2615..3649
    /gene="hemZ"
    2615..3649
    /gene="hemZ"
    /note="hemZ"
    /note="Rv1485, (MTCY277.06), hemZ, len: 344. Ferrochelatase similar to e.g. HEM2_BACSU P32396 ferrochelatase (310 aa), FASTA scores, opt:490, E(): 2e-24, (30.2% identity in 295 aa overlap)"
    /codon_start=1
    /transl_table=11
    /product="hemZ"
    /protein_id="CAB02035.1"
    /db_xref="GI:1524231"
    /db_xref="SWISS-PROT:P71765"
    /translation="MQFDVLLLSFGGPEGPQVRPFLNVTTRGQVPAERLDAVAEHLHFGGVSPINGINRLLIAELEAQQLPYVPCGNRMPEYVEDAVTAMDNQVRAAVFAVSAMGYSCTQYVEDIARARAAAGRADAPLVKLPYFDHPLFVEMFADAITAAATVRGDARLVFTAHSTPTAADRRCGNLYSRQVAYATRLVAAAAGYCDPDLAQMSRSGPQVPLPDPVTDQLTGLAGAGINAVIVCPGVADHIEVWDLDELRQDAEAGIAYARASTPNADPRFARLARGLIDELRYGRIPARVSGPDPVPGCLSSINGQPCRPPhCVASVSPARPSAGSP"
    complement(3615..4523)
    /gene="Rv1486c"
    complement(3615..4523)
    /gene="Rv1486c"
    /note="Rv1486c, (MTCY277.07c), len: 302. Highly similar to M. avium YS07_MYCAV 007402 hypothetical 33.5 kd protein mav321 (320 aa), fasta scores; opt: 1217 z-score: 1502.0 E(): 0, 71.1% identity in 315 aa overlap"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein Rv1486c"
    /protein_id="CAB02036.1"
    /db_xref="GI:1524232"
    /db_xref="SWISS-PROT:P71766"
    /translation="MIRVSRVEQLHCKIMQPSVSLSIWANAWLAGKAAPDDVLALSIMAPTQSVAAVDAVAAAGHTGLPMPDVGAGTVSLQTLRAAVGRRRLRGTINVLVPGDVVRLAAGTQFEHDALALAGEAVIVANPEDGSAVGLVPEFSYGDVDEAAQSEPLTPELCALSMWVSLPGAPVLEHYELGDAEYALRSVAGSAEALSTIGLSSDVAKPRGLVEQLLESSRQHRVDPDHAPSRALRVLENAAHDAIIAVSAGLSRLPIGTQSLSDAQRATDALRPLTAVVRSARMSAVTAIILHSAMPD"
    4539..4973
    /gene="Rv1487"
    4539..4973
    /gene="Rv1487"
    /note="Rv1487, (MTCY277.08), len: 144. Highly similar to M. avium TR:O07404 (EMBL:AF002133) MAV145. (145 aa), fasta scores; opt: 667 z-score: 1148.3 E(): 0, 72.5% identity in142 aa overlap"
    /codon_start=1
```


Db	9698	YVDAYS	CSRYDARC	YDACYDS	YSNSTCYD	ATBT	TSRYS	STYS	SYSN	CYDAT	T	SRC	T	BY	STB	96399
QY	1104	ggttag	ggcgga	cta	ctg	gcttg	acg	gccttc	gcttc	gtc	gtc	gtc	gtc	gtc	gtc	11633
Db	9638	TBT	TSRCAK	CTBD	ST	STAK	NST	ST	ST	TRCT	BY	SSR	SG	SYCS	RSR	95799
QY	1164	gacg	gcg	gtgcg	gag	tgag	ccga	ctgcg	gcgcg	atcg	caat	cccc	aagc	ctt	gtg	12233
Db	9578	CYST	TYST	TYA	SC	TS	SR	GS	YD	AS	R	ST	ST	SR	CT	95199
QY	1224	gatg	gcta	gcttg	gcgcg	gttg	gcag	aatg	ctg	tg	gcag	g	ctc	gtt	gata	12833
Db	9518	YSSD	SDCT	BYSS	DRCS	RS	DS	ST	CN	CYS	CS	DS	R	YST	TYD	94599
QY	1284	ctat	cgac	cgga	agcc	a	ga	ctc	gtc	gtc	gtc	gtc	gtc	gtc	gtc	13433
Db	9458	SR	CYT	N	ST	YS	CS	R	T	B	S	R	N	S	C	93999
QY	1344	gcag	cccc	aa	cc	ctt	cc	ca	ag	ca	act	c	cc	ct	ca	1374
Db	9398	TBY	S	Y	S	C	T	BY	S	N	S	T	Y	A	S	9368

```

complement(623. ,1075)
/gene="m112104"
/note="unknown protein"
/codon_start=1
/translation_table=11
/protein_id="BAB49317.1"
/db_xref="GI:14022709"
/translation="MSWNRAGSEGEIKRLPGRPKSLIVCGNDITFVVRVYGNRRPP
CGPRRASSPLRLRLAGVLFAIQKTEIPGTRAAYGMQELGQATAQQLTHGGRARRR
ALVETPCYEGREFFLVQHLLQPFRTLEIVSHRGPFLLPMQPVSRMSK"
1184. .1534
/gene="mlr2105"
1184. .1534
/gene="mlr2105"
/note="unknown protein"
/codon_start=1
/translation_table=11
/protein_id="BAB49318.1"
/db_xref="GI:14022710"
/translation="MDRPLYALAEFFSKPPGFYVMTAALVCAALLTSATLGTIISIS
ALCTIGVVLIONRYDTAAMQAKNELIVALDTARNEVVGLEHASPKAIDAELEKNENR
ASTAPVETETODPA"

```

RESULT	15
AP002999/c	
LOCUS	
DEFINITION	AP002999 346294 bp DNA
ACCESSION	Mesorchizobium loti DNA, complete genome, section 6/21.
VERSION	AP002999 BA000012
KEYWORDS	AP002999.2 GI:14022707
SOURCE	Mesorchizobium loti (strain:MAFF303099) DNA.
ORGANISM	Mesorchizobium loti

```

gene      complement(1531. .2097)
          /gene="m112106"
          /complement(1531. .2097)
          /gene="m112106"
          /note="hypothetical protein"
          /codon_start=1
          /transl_table=11
          /protein_id="BAB49319.1"
          /db_xref="GI:14022711"
          /translation="MIDCDARVARSIRFLSCGSRACKFSPNGSQNARREEMPRREGCE

```

REFERENCE
1 (sites)
AUTHORS
Kaneko, T., Nakamura, Y., Sato, S., Asamizu, E., Kato, T., Sasamoto, S.

gene

TITLE	INVENTOR(S)
Complete genome structure of the nitrogen-fixing symbiotic	20030002/001 Yamada, M. and Iadad, S.

JOURNAL DNA Res. 7 (6), 331-338 (2000)
MEDLINE 21082930
REFERENCE 2 (bases 1 to 346294)
AUTHORS Kaneko, T.
TITLE Direct Submission
JOURNAL Submitted (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research

gene

URL: <http://www.kazusa.or.jp/rhizobase/>,
Tel: 81-438-52-3935 (ex. 2338), Fax: 81-438-52-3934
On May 11, 2001 this sequence version replaced qi:11994974.

nos

gene

gene

CDS

gene complement(623.1075)
/gene="m112104"

အခြေခံ

```

complement(623. .1075)
/gene="ml12104"
/note="unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB49317.1"
/db_xref="GI:14022709"
/translation="MSWNKAGSEGEIKRLPGRPKSLIVOCNDITFVVVRVGNRRP
CGPRASSPLRLRLAGVLFATIQKTEIPGTRAAVGMQEGALGGQATAOQLTHGGRARR
ALVETPCVEGREFFLVQHLPFRTLEIVSHRGCPFLPAMQPVSRMSK"
1184. .1534
/gene="mlr2105"
1184. .1534
/gene="mlr2105"
/note="unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB49318.1"
/db_xref="GI:14022710"
/translation="MDRPLALAEFFSKPPGFYVMAALVCALLTSATIGTYIISIS
ALCLTGVVLIONRDPAMQAKLNEIIVALDTRNEVYGLEHASPKAIDAEKVENRR
ASTAPAHETQPPA"
complement(1531. .2097)
/gene="ml12106"
complement(1531. .2097)
/gene="ml12106"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB49319.1"
/db_xref="GI:14022711"
/translation="MIDCDARVARSIRFLSCGRKAGFSFNGSQNARREMPRECE
GENLGTENRHGSCGPPTFESIFMEKTFNRMAQVAHATGRPWAFALCTASYLWAAT
GPVFHYSETWQVLINTGTITITFLMFLIQNTQNRDGAIOVKLDELIRSGIKNDPI
GIEHLTEREVEAFRRACEAKTAKASKG"
complement(2346. .2585)
/gene="ms12108"
complement(2346. .2585)
/gene="ms12108"
/note="unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB49320.1"
/db_xref="GI:14022712"
/translation="MVSMEFGAQSVFLGPEDLAMCQRIFFQVRADANWERSAVESEE
LAASVILAFQHGFTDEALLAEIRSQDRRLMQRTG"
complement(2682. .3116)
/gene="ml12109"
complement(2682. .3116)
/gene="ml12109"
/note="unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB49321.1"
/db_xref="GI:14022713"
/translation="MTAFRARPRSPLTIRKYSTTLGWEDGLLAGTTMKRRORSPPRID
SPFLYESVHSRBRPSTAYGRLVANGAPAIYAETEMAEHSDSKVTYALDCEILRGAF
RKSVEHRYAEREMRMHARVLARELTETDEIDPDILDWIVRK"
3560. .4291
/gene="mlr2110"
3560. .4291
/gene="mlr2110"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB49322.1"
/db_xref="GI:14022714"
/translation="MSDPTWDGSPQYRNELLREMPADDIALALPHNQRCALPVRMMLV
TPNPIPEAVYFIEQIGSVAVSTASGHDAEIGFEGEGMTGSALVMGDDRAHACFVQ
LEGEAIRDAAPFNALAAASPTLRLELLHFVNTLHTQTGCTALVNARKLEERLARWL
LMCDDRPGESLMTHEFLSIMLGYRRPGVTVALQLLEGRLIRSRGEIIVIRDRAGL
LELANGYGOAEAYARLIGEIMPG"
5065. .5808

```

```
CDS      /gene="mlr2113"
5065. .5808
/gene="mlr2113"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB49323.1"
/db_xref="GI:14022715"
/translation="MPIRKNLAANLRLLVSGHASVAAVCRGLNMNRTQFERYLQGSV
PNRATARLICTYFRIDEDDLVRAPEVPAPEKALLPIHOTLYENMVGPAPAVAGTY
FTYFTVPDRDILLVRSVTFVRREALVTFRRVTRWAEGRQGGARVLGWHYCAISR
LNMIFYAGINRQTRPESVMTYQWAPFSDPVLVGKAFVLGTSGPACVSVIMROEVGRIS
LRQVLGMCRRVSLDDPRLDQLVASLVREG"
5859. .6530
/gene="mlr2114"
5859. .6530
/gene="mlr2114"
/codon_start=1
/transl_table=11
/product="probable transcriptional regulator"
/protein_id="BAB49324.1"
/db_xref="GI:14022716"
/translation="MLNSFDYGGKEADTVHALEEDIIIFRLAPGTRLVEDVLLARFPV
SRHTIRQALYQLERLGIIVTRERNKGANVRRLSPPEVRQIYEVREMIQROALMIPLV
SEALIAOLEEIHRIYSGHVADLHEANDREHLLTMFACGNAYLVSSIHYMRLS
LPVRANSLADREKLDVSRQHHAMMIEALKRRDNWVLAQLCVDBIQPSKVIFYLOEIEIETA
PATG"
complement(6597. .7601)
/gene="ml12116"
complement(6597. .7601)
/gene="ml12116"
/codon_start=1
/transl_table=11
/product="probable zinc-binding oxidoreductase"
/protein_id="BAB49325.1"
/db_xref="GI:14022717"
/translation="MMKAFVVDKYQKKALRLAEMPEPELRDNDVVEIHAAGVNLDD
SKLRDGEFKPILPYRPEILGHDVAGIVRAGSKARKFKLGDEVYARSDGRIGTFAQ
SIAMDEADVALKPKNLMEEAASIPLVGLTAMQVLVETAGLKKQKVEIQAGSGVGT
FAIQLAHGLGATVATASANADLVKGLGADIVIDYKKDDEKILQGYDVLNSQDAK
TLEKSLHVLKPGKLLISGPPDPDFAGOKLNMVLKLVRLSLRGIKSKARRAGVG
SFLFMWAOGDQIGEITSRIESGFIRPIDRIFPEKTNEALAYEAGRAKGVVVKVR"
complement(7723. .8538)
/gene="ml12118"
complement(7723. .8538)
/gene="ml12118"
/codon_start=1
/transl_table=11
/product="oxidoreductase"
/protein_id="BAB49326.1"
/db_xref="GI:14022718"
/translation="MNRNSGKTATVGTGASSGIGRASAEILARAGTVFGTSRKASQGS
SNGISMLVCDVDDASVASLSTVLGTRIDLLVNNAGIGLGAEESSVSQOALF
DVNLFGVMKTNVAVLPSMRRLGGRIVNIGSILGLVPAPYSAHYSAYKHALLEGYSESL
DHEVRAFNIRVSVLEPAFVRTVFQNGFEPDSVMKVYDEARAGLEALLRDVMPKADLP
EVVAKVVLRAATDRHPLRRTYTAGKAARQVSMRLRFAPTGMFDKILRKQERLPV"
8432. .8617
/gene="msr2119"
8432. .8617
/gene="msr2119"
/note="unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB49327.1"
/db_xref="GI:14022719"
/translation="MPKTVKPARAKISALARPPEAPVTIAVFPERRFIVKAPNVRG
LACDRRLNYDRNITLCS"
8646. .9233
/gene="mlr2120"
8646. .9233
/gene="mlr2120"
/codon_start=1
```

```
/transl_table=11
/product="probable transcriptional regulator"
/protein_id="BAB49328.1"
/db_xref="GI:14022720"
/translation="MRYDKGRKSDASRSRIMEVASHREFRGDGIASGASIMSDAGMTN
GAFYPHOSKADLVRESMASALETOSOOLOALASGGLLAMAMYLSPEDRDHPOTGC
ASALLPEIARQPPETRSLYTEKLLTLVHQLAKDLPPTRPDEGVALGVFAALIGALQ

Query Match      3.0%; Score 42.8; DB 1; Length 346294;
Best Local Similarity 56.3%; Pred. No. 4.7;
Matches 80; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 1055 agcgcgcttaagcggttgagcgccctaccgtcccgctggaagtcggttgaggcg 1114
    |||  ||  |  |||  ||  |||  |  ||  |||  |||  |||  |||  |||  |||  |||
DB 7925 AGGCATTGCTTCCGACGCGATGATGCCCAAGGCCGATCTGCCCGAAGTGCGGAAGTGG 7866

QY 1115 gactactgggcttaagcgccctcgcttgctgtgctgtgtacagcggtgacggcggtgc 1174
    ||  ||  ||  |||  ||  |  |||  ||  |||  ||  |||  ||  |||  ||  |||
DB 7865 TTCTGCGGGCTGCGACCGACCGGCATCCGCTGCGGCGTACACCGCGCAAGCAGCGC 7806

QY 1175 ggcaggtgagcgactgcggcg 1196
    |||  |||  |||  |||  |||  |||
DB 7805 GGCAGGTAAGCATGCTGCGCGG 7784
```

Search completed: January 10, 2002, 22:26:09
Job time: 3192 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 10, 2002, 21:01:57 ; Search time 1154.57 Seconds
(without alignments)
13067.266 Million cell updates/sec

Title: US-09-887-038-2
Perfect score: 1404
Sequence: 1 atgactgtctgcaactct.....atcagacgaaaaaatgtag 1404

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_etc:*
10: gb_est1:*
11: gb_est2:*
12: gb_hic:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_vrt:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	52.4	3.7	925	13	CNS0091P
2	47.2	3.4	1101	13	CNS017SY
3	46.2	3.3	500	10	BE602908
4	43.8	3.1	645	13	CNS01213
5	43.4	3.1	922	13	CNS0073W
6	42.4	3.0	604	13	AZ935264
7	42.4	3.0	925	13	CNS0091P
8	42.4	3.0	1201	13	CNS016BO
9	42.2	3.0	970	13	CNS010C9
10	42.2	3.0	997	13	CNS006DN
11	42	3.0	1250	10	BE545125
12	41.8	3.0	846	13	CNS010RJ

C 13	41.8	3.0	1201	13	CNS01679	AL106383	Drosophila
C 14	41.8	3.0	1868	11	BG167536	BG167536	602342745
C 15	41.6	3.0	1100	13	CNS016KD	AL106855	Drosophila
C 16	41.2	2.9	1101	13	CNS012S8	AL101954	Drosophila
C 17	41.1	2.9	446	11	BG274660	BG274660	WHE2237_E
C 18	40.8	2.9	464	10	AI858018	AI858018	wj69c12.x
C 19	40.8	2.9	663	13	CNS032G8	AL224801	Tetraodon
C 20	40.6	2.9	912	13	CNS006N3	AL065775	Drosophila
C 21	40.2	2.9	914	13	CNS00CZP	AL059740	Drosophila
C 22	40	2.8	339	10	AW582345	AW582345	MR3-ST022
C 23	40	2.8	827	11	BG845859	BG845859	1024011E0
C 24	39.8	2.8	844	13	CNS0052P	AL056552	Drosophila
C 25	39.8	2.8	977	13	CNS014H2	AL104144	Drosophila
C 26	39.8	2.8	1101	13	CNS00LXJ	AL078875	Drosophila
C 27	39.6	2.8	910	13	CNS006ON	AL065629	Drosophila
C 28	39.6	2.8	937	13	CNS006XP	AL066056	Drosophila
C 29	39.2	2.8	764	11	BI151035	BI151035	602916857
C 30	39.2	2.8	1159	13	CNS015XR	AL106041	Drosophila
C 31	39	2.8	442	10	BE725159	BE725159	894081D05
C 32	39	2.8	839	13	CNS004NB	AL054280	Drosophila
C 33	38.8	2.8	550	11	BI181863	BI181863	UNL-P-FN-
C 34	38.8	2.8	1101	13	CNS016HG	AL106750	Drosophila
C 35	38.6	2.7	1101	13	CNS00EOE	AL067626	Drosophila
C 36	38.6	2.7	1101	13	CNS00EA5	AL068833	Drosophila
C 37	38.4	2.7	305	10	AV631321	AV631321	AV631321
C 38	38.4	2.7	407	11	BG908379	BG908379	TaLr1167F
C 39	38.2	2.7	431	10	AL562202	AL562202	AL562202
C 40	38.2	2.7	972	13	CNS02BJJ	AL189928	Tetraodon
C 41	38	2.7	1101	13	CNS014Y2	AL104756	Drosophila
C 42	37.8	2.7	765	11	BF865981	BF865981	963063E12
C 43	37.6	2.7	321	11	D24188	D24188	RICR1503A R
C 44	37.6	2.7	472	11	BG464251	BG464251	EM1_71_E0
C 45	37.6	2.7	681	10	AI403917	AI403917	GH23463.5

ALIGNMENTS

RESULT 1	CNS0091P	925 bp	DNA	GSS	03-JUN-1999
LOCUS	Drosophila melanogaster genome survey sequence TERT3 end of BAC #				
DEFINITION	BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.				
ACCESSION	AL053013	GI:4934461			
VERSION	AL053013.1				
KEYWORDS	GSS.				
SOURCE	fruit fly.				
ORGANISM	Drosophila melanogaster				
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
REFERENCE	1 (bases 1 to 925)				
AUTHORS	Genoscope.				
TITLE	Direct Submission				
JOURNAL	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr				

COMMENT

- Web : www.genoscope.cns.fr
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.


```

FEATURES                                  Location/Qualifiers
      source                               I . 925
                                           /organism="Drosophila melanogaster"
                                           /db_xref="taxon:7227"
                                           /clone_1ib="RPCI-98"
                                           /clone="BACRI9D16"
                                           /note="end : TER3"

BASE COUNT    120 a        61 c         61 g       172 t     511 others
ORIGIN

Query Match               3.7%; Score 52.4; DB 13; Length 925;
Best Local Similarity   15.2%; Pred. No. 0.012;
Matches   57; Conservative 160; Mismatches 159; Indels   0; Gaps   0;

OY    759 ttgttgtggcgcttataggcgtctaactggttcaaccccgtaaccgcagccctggcg 818
      |||||:::||::||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db    549 TTTTSSSGSYGKGCGSSGSGBSCGCCSSCSGCCSSSSCSCCBCCCCCCSSSYCCSSSBSS 608

OY    819 acgcgtgcattcccagtcgatgttggtgtagctagtgcggtgctctttgttgccgtgct 878
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db    609 KCSSTBSBCSCCCSSKSVCGTCSSSSSSCSSSSSTSSSTSKSSSSGGSSSSSSSS 668

OY    879 tggacttgagccgltgcgcgtgcgctgttagcatcttgttgggcggtgaagacagcag 938
      ::||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db    669 YTTAKSTAASGSGSWAGGSGGTGSTSSSSSSSSSTSTSSSVSSGSKSTBBSGGBSS 728

OY    939 caacaattccgataatgtctgtgctggcgtgtgctgcagatatcataaatcgccttg 998
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db    729 GSSSSSSSTSBBCTSTSSSSSSSYSSSTGCSCCTCCSYSSYSTSSSTSMWGSTRGSS 788

OY    999 gctggcatalcgcgccgcgcataaccgctttaacctggttacccctcatacacagc 1058
      ::||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db    789 SSSVGTSSSDSTSTCCSCCYMCCTCTYBMBCYTSTSCGSSSSSGKGVTKCGCGGC 848

OY    1059 gcgcttacggcgtltagcgctaacctccgtcccgctggaagtgcggttagggcgact 1118
      ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db    849 GSSSTNGMBGTSSACSSSSSSSSSVSSSSSKSSASSSVSSGCVSSNSSASAKSSS 908

OY    1119 actgggcttgacggc 1134
      :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db    909 SCGYSSGSGSGSGSVS 924


RESULT 2                                CNS017SY          26-JUL-1999
CNS017SY/C                             Drosophila              GSS
LOCUS                                   Drosophila melanogaster genome survey sequence SP6 end of BAC
DEFINITION                            BACN37L08 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION                              AL108460
VERSION                                AL108460.1 GI:5628764
KEYWORDS                                fruit fly.
SOURCE                                 Plasmid Drosophila melanogaster
ORGANISM                               Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephyrdoidea; Drosophilidae; Drosophila.
REFERENCE                               Genoscope.
AUTHORS                               Direct Submission
TITLE                                 Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr - Web : www.genoscope.cns.fr)
JOURNAL                                Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobACL1.
COMMENT                                location/qualifiers
```

source 1..1101
/organism="Drosophila melanogaster"
/plasmid="pBelobAC11"
/db_xref="taxon:7227"
/clone_1lb="DrosBAC"
/clone="BACN37L08"
/note="end : SP6"

BASE COUNT 254 a 176 c 160 g 152 t 359 others

ORIGIN

Query Match 3.4%; Score 47.2; DB 13; Length 1101;
Best Local Similarity 17.7%; Pred. No. 0.25;
Matches 80; Conservative 171; Mismatches 201; Indels 0; Gaps 0;

QY 504 cggcgttgaagagctgcgacttggytgcacatcgcactcgtttgcgacttcacctcacg 563
::: | : : : | : : : : : : : : : | : : : : : : :
Db 1038 SSBSSSTSSSSSSSSSSSTSSSSTSBHPTSTSSSTTTTITTTTTTWTBTBBSSBSBS 979
::: | : : : : : : : : : : : : : : : : : : : : : : :
QY 564 gglttacagtattctggcaaccaccaactcgtcgtcgtcltatctlygtccgaactgc 623
: | : : : : : : : : : : : : : : : : : : : : : : :
Db 978 SSSSTSSSBTKSBTSBSBSTSSSAHBSSSSSSSSBSSTSTBTBSTSHBSTSSSSSGSSS 919
: | : : : : : : : : : : : : : : : : : : : : : : :
QY 624 ctlttcgcagcagcagcatcgggtgtgycgcgcgtgcgtccccaaactcgtgcgatcgc 683
: : : : : : : : : : : : : : : : : : : : : : : : :
Db 918 SSBBSSTBSBSBTBTBTBTBKSTSTSTSSSTSBTTTBSBSSSSSSSSBSSTTSBS 859
: : : : : : : : : : : : : : : : : : : : : : : : :
QY 684 tgcagacaggtgcgagcagcttatgtctgatacctaactaacagtcgcgtgcgtgcg 743
| : : : : : : : : : : : : : : : : : : : : : : : : :
Db 858 TBSSTGSSSBCTCTSKCSTBTSGSTPRTBMSKBSSTSTSTSTSTGSSBGGSCGSTSGSSC 799
| : : : : : : : : : : : : : : : : : : : : : : : : :
QY 744 ttgtgcgccatgatttctgtcgcgtlatalagycgtcactggtttaaccocgtct 803
: : : : : : : : : : : : : : : : : : : : : : : : :
Db 798 SSBGKSTSSSGSTSSGTTCGBCSSTGGCCSCCCYCSCCCTCGTGTSTGCCSCSCSGTGSSB 739
: : : : : : : : : : : : : : : : : : : : : : : : :
QY 804 acccgaccctgcgcagcgtgctatcccagtcglatgttgttgtaactagtcgcgtgct 863
| : : : : : : : : : : : : : : : : : : : : : : : : :
Db 738 KCCVSTSSASASTGSSCSSSTSSSGGKMTGGCCGHTTKGTGTFCCCTGGBYCCYBCMT 679
: : : : : : : : : : : : : : : : : : : : : : : : :
QY 864 ctgtgtgcgctgtctgtgacttgagccggttgcgcgtgcgcgtgttgagcatcttgtgg 923
: | : : : : : : : : : : : : : : : : : : : : : : :
Db 678 MMCCGSSSCSCSYCCGSSSKCTSMGTTGCMYYCCGCTSCCGCTGGGGMASACGTGCTGS 619
||| : : : ||| : | | || : | : |
Db 618 GCGNMVMGMAGAVTATAANTTTNTNAAMAA 587
||| : : : ||| : | | || : | : |

RESULT 3
BE602908 500 bp mRNA EST 02-MAR-2001
LOCUS HVSMEH0100N13f Hordeum vulgare 5-45 DAP spike EST library
DEFINITION HYCDNA0009 (5 to 45 DAP) Hordeum vulgare CDNA clone HVSMEH0100N13f,
mRNA sequence.
ACCESSION BE602908
VERSION BE602908.2 GI:13190663
KEYWORDS EST.
SOURCE barley.
ORGANISM Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 500)
Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Fritsch,D., Yu
X., Anderson,H., Dale,J., Henry,D., Kernodle,S., Palmer,M., Rambo
J.T., Saski,C., Schwartzbeck,J., Simmons,J., Choi,D.W., Main,D. and
Wood,T.
Development of a genetically and physically anchored EST resource
for barley genomics
Unpublished (2000)
On Aug 21, 2000 this sequence version replaced gi:9860469.
COMMENT JOURNAL
Contact: Wing RA
Clemson University Genomics Institute

Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: AATTAAACCTCATAAGG
High quality sequence start: 3
High quality sequence stop: 472.

FEATURES

Location/Qualifiers
1..500

/organism="Hordeum vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HVSMEh0100N13f"
/clone_lib="Hordeum vulgare 5-45 DAP spike EST library
HVCDA0009 (5 to 45 DAP)"
/tissue_type="5-45 DAP Spike"
/lab_host="SOLR"
/note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI; For
more details on library preparation and sequence analysis
see http://www.genome.clemson.edu/projects/barley/ To
order a clone see http://www.genome.clemson.edu/orders"

BASE COUNT

97 a 137 c 162 g 103 t 1 others

Query Match 3.3%; Score 46.2; DB 10; Length 500;
Best Local Similarity 53.3%; Pred. No. 0.35;
Matches 96; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

OY 1057 ggcgccttaagcgcgttgagcgcctactccgcgcgtggaagtcgcggttgagggcga 1116
||| | ||||| | | | ||||| | | | ||||| | | | ||||| | | | ||||| |
Db 72 GCGGCCCGCCGCGGTGCAACGGCGGACCGCCCGCGGACGAGGGGTGAGAGAGAGC 131
OY 1117 ctactggccttgacgccttcgcctgtgctgctgtgtcaagcgcgtgacgcgcgtgcg 1176
|| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 132 CTGAGGTCCTCNGCCGCGTCCGCTCTGGGGCGGCATCTGCTCGAGGTGGGACG 191
OY 1177 cagtgagccgactgcgcgcgatcgcaatcccaagcctttgtgtgatgtagcttg 1236
| |||| | | | | ||||| | ||||| | | | | | | | | | | | | | | | | | | |
Db 192 CGCGTGATGTCCGCTGGCGGACAGAAATCCACCCCGTTAAGCATCATCGAGCGCCTG 251

RESULT 4

CNS01213

LOCUS 645 bp DNA GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC
BACN08C07 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.

ACCESSION

AL101589 GI:5613200

VERSION

KEYWORDS

SOURCE

ORGANISM

fruit fly.
Plasmod Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 645)

REFERENCE

Genoscope.
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)

TITLE

JOURNAL

COMMENT

Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelobAC11.

FEATURES

Location/Qualifiers
1..645

/organism="Drosophila melanogaster"
/plasmid="pBelobAC11"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"
/clone="BACN08C07"
/note="end : T7"

BASE COUNT 28 a 26 c 85 g 92 t 414 others
ORIGIN

Query Match 3.1%; Score 43.8; DB 13; Length 645;
Best Local Similarity 11.6%; Pred. No. 1.5;
Matches 57; Conservative 123; Mismatches 311; Indels 0; Gaps 0;

OY 760 ttgtctggcgttattagcgtctactgtgttcaacccgtctaacccgacccctggcga 819
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 98 TCKKCGCNGNNNNNNNNNGGCCNNNTNGKKKGNGGNGNGTGTGTGNGNGGGGGG 157
OY 820 cgtcgtatccacgtcgtatgggtgactagtcgcgtgctctgtgtgctgctt 879
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 158 GGNNNNGCANNNTTNTNGCTTNNNGCANNNNNSCCSCMCNTCANNNTTNTNSSSSSSNS 217
OY 880 ggaactgagccgttgcgctgctgctgtgagcatcttggggcggtgaagacagcagc 939
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | |
Db 218 NNNNTNNNNNSNNNSNNNSNNSSSSNNNNNSGGSSSSNSNSNNSSNSNTSSN 277
OY 940 aacaactccgcatcaatgtctgctgctgcgtgtgtcagatgatcaagatcgcccttg 999
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | |
Db 278 NNSNNTNSNNNSNNNSNNNSSTNSNNSSSSTSTNTNTSNTSSNSNSATGSSSSNSSSGGT 337
OY 1000 ctggagcatcgcccgcaataacgccttcaacctgttatccctcatcaagcgcg 1059
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | |
Db 338 TBSGSSSSSSSSNSNNNNNNNSBTSNSNNSGATTSSTNGSSSSSSSSSSSTSS 397
OY 1060 cgtttacggcgttgagcgcctactccgtccgcgtggaagtcgcgttgtaggcgagcta 1119
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | |
Db 398 NTSTSTSTSSNSNTSTSTSSSTSTNTSTSTSSSSNSSTGSSSTTTTNSASSNSNNN 457
OY 1120 ctggccttgacgccttcgttgctgctgctgctgacgcgcgtgacgcgcgtgcgag 1179
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | |
Db 458 NGNSSSMAANSTSTTSSNTTSSNNSSCNTTSANNTSSTNTTNTAGNNNNATSSSN 517
OY 1180 gtgagccgactgcgcgcgatcgcaatcccaagccttgggttgatgtagctagctggcc 1239
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | |
Db 518 SNNNSNNNSNNNSNGSGTSSGNTSSSCTGTBSSGNSNNNSCNGSNNNSCTGSSN 577
OY 1240 ggttgagcag 1250
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | |
Db 578 SSSTTTGSSSS 588

RESULT 5

CNS0073W

LOCUS 922 bp DNA GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TERT end of BAC #
BACR14D09 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.

ACCESSION

AL066784 GI:4945247

VERSION

KEYWORDS

SOURCE

ORGANISM

fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 922)

REFERENCE

Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).

The BDGP is constructing a physical map of the *Drosophila melanogaster* genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP *Drosophila melanogaster* BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCT-98 and was constructed by partial EcoRI digestion of *Drosophila* DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES	Location/Qualifiers
source	1. .922
	/organism="Drosophila melanogaster"
	/db_xref="taxon:7227"
	/clone_lib="RPCI-98"
	/clone="BACRL4D09"
	/note="end : TET3"
BASE COUNT	223 a 95 c 109 g 221 t 274 others
ORIGIN	

Query Match	3.1%;	Score 43.4;	DB 13;	Length 922;
Best Local Similarity	18.8%;	Pred. No. 2.1;		
Matches	72;	Conservative 131;	Mismatches 179;	Indels 0;
				Gaps 0;

[illegible]

RESULT	6
AZ935264/C	
LOCUS	
DEFINITION	AZ935264 604 bp DNA GSS 24-APR-2001
ACCESSION	BJ_Ba0003H07f B. japonicum BAC library Bradyrhizobium japonicum
VERSION	AZ935264
KEYWORDS	genomic, DNA sequence.
SOURCE	AZ935264.1 GI:13777704
ORGANISM	GSS.
REFERENCE	Bradyrhizobium japonicum.
AUTHORS	Bradyrhizobium japonicum Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group; Bradyrhizobium group; Bradyrhizobium.
TITLE	1 (bases 1 to 604) Tomkins, J.P., Wood, T.C., Stacey, M.G., Ioh, J.T., Judd, A., Golcochea, J.L., Stacey, G., Sadowsky, M.J. and Wang, R.A. A marker-dense, sequence-ready map of the Bradyrhizobium japonicum

genome
Genome Res. 11 (8), 1434-1440 (2001)
MEDLINE
21376150
COMMENT
Contact: wing RA

Tel: 864 656 7288
Fax: 864 656 4293
Email: rwinge@clemsson.edu
Class: BAC ends
High quality sequence stop: 548.

```

FEATURES
source
Location/Qualifiers
1.604
/organism="Bradyrhizobium japonicum"
/strain="USDA110"
/db_xref="taxon:375"
/clone_lib="B. japonicum BAC library"
/lab_host="E. coli"
/note="Vector: pindigo536; Site_1: HindIII"
BASE COUNT
90 a 197 c 219 g 98 t
ORIGIN

```

Query Match	3.0%;	Score 42.4;	DB 13;	Length 604;
Best Local Similarity	54.5%;	Pred. No. 3.3;		
Matches 85; Conservative	0;	Mismatches 71;	Indels 0;	Gaps 0;

QY	1054	caggcgcgcttaagcggttagcgccactaccgcgcggaagtcgcggttagggc	1113
Db	545	CTGGCTCACACACAGCGCCCTTGCCGCCGGCCTCAATCTGCGGAGTTGACGCGCAAGGAT	486
QY	1114	ggactactggtgcttagcgcccttcgcttgctgtgtgtcaacgycggtgaagcggtg	1173
Db	485	ATCCTGAAGGAATTGACCGCGCCTGGCGCCGACAGCTGCTGCCATATGCGGAGACGGTGTGG	426
QY	1174	cggcaggtgagccgactgcgycgatcgcaatccc	1209
Db	425	CGGCTGCTCGACATCAAGCGCCGCGAAGGCAAGCGC	390

RESULT 7
CNS0091P/c
LOCUS
DEFINITION
CNS0091P 925 bp DNA GSS 03-JUN-1999
Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR19D16 of RCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
AL053013
AL053013.1 GI:4934461
GSS.
fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 925)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the *Drosophila melanogaster* genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP *Drosophila melanogaster* BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of *Drosophila* DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the library P1 and EST libraries. A more detailed description of the library

and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source

location/Qualifiers
1. .925
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPC1-98"
/clone="BACR19D16"
/note="end : TET3"

BASE COUNT 120 a 61 c 61 g 172 t 511 others
ORIGIN

Query Match 3.0%; Score 42.4; DB 13; Length 925;

Best Local Similarity 14.3%; Pred. No. 3.7;

Matches 52; Conservative 163; Mismatches 145; Indels 4; Gaps 1;

QY 842 tgggtgagactgcggtgctctgtgtggcggtgtgacttgagccgttgcggtgc 901

Db 912 BSCSSSSMSTSSSNSBSCSSSSSBSSSTSSMSSSSSBSSSSSSSGTSSACVKCNA 853

QY 902 gcgtgtgagcatcttgttggggcggtgaagacagacacacacttcgcatcatgtct 961

Db 852 SSSCGCCGCMABCCMCCSSSSSCCG---SASARGVKVRASGAGAKRGGGGSASASHSS 797

QY 962 ggctggcggtgtcagatgatcaagatcgcccttgctggcgtgcgtcccgcaata 1021

Db 796 SSACBSSSSSCASCMASASSSSASSSRSRSGGAGGSGAGSSSSSSSSASAGSVSS 737

QY 1022 ccgccttaacctgtttatccctctatcaacagcgcgcttaccggcttgagcgct 1081

Db 736 ASSSSSSSSSVSCSVASSMSCSBSBSASASSSSSSASACASCCCTTSMSCSCT 677

QY 1082 actccgtccgctggaagtcgcggttgaggcggaactacttggttgacggccttcgtt 1141

Db 676 SASMSAARSSSSSSSCSSSMASASSSSSSSGSSSSSGSACBSMSSGGGSG 617

QY 1142 ggctgctgtgtcagcggtgacggcggtgcggtgagccgactgcggtgcgcatc 1201

Db 616 SVSASSGMSVSSSGGRSSSGGGGCGVSGSSSSSGSGSGSVCSGCMCRCSG 557

QY 1202 gcaa 1205

Db 556 SSAA 553

RESULT 8

CNS016BO/c

LOCUS

DEFINITION

CNS016BO 1201 bp DNA GSS 26-JUL-1999
Drosophila melanogaster genome survey sequence SP6 end of BAC
BACN15M12 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
<http://www.edgp.ebi.ac.uk> - This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton

and Genevieve Payan. It has been constructed in the vector
pBelOBAC11.

FEATURES

source

location/Qualifiers
1. .1201
/organism="Drosophila melanogaster"
/plasmid="pBelOBAC11"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"
/clone="BACN15M12"
/note="end : SP6"

BASE COUNT 292 a 254 c 222 g 205 t 228 others
ORIGIN

Query Match 3.0%; Score 42.4; DB 13; Length 1201;

Best Local Similarity 29.4%; Pred. No. 4;

Matches 87; Conservative 73; Mismatches 136; Indels 0; Gaps 0;

QY 587 ccaactgctgctgtctatctgtgtgcgacgactgctcttctgcagcagcagcgagg 646

Db 1186 CCCYCCCCCCCCCTCTYTCGGBYCBYBSYCYSCYTTTBYTCGGBSBSGGG 1127

QY 647 tgtgcgcgctggtctcccaagctgtgcgcatcgctgcgacaggtgcagcgttat 706

Db 1126 KGGCCSSYSBGSBGKKGCCGCTKSRTCTGTTTGKYGSGGGSGSGGMVTCSCCC 1067

QY 707 gtctgatccctacctaagtcgcggtgtgctgtgtgttttcgccatgattttgtct 766

Db 1066 CGCMSCCTGBCCSCCSCGSKBGGKGGGGCGGCGCKTHTSCGGGGGGGTTKSKGVT 1007

QY 767 gggcggtataggctctactgtttcaaccccgctcaccgacacctggcgagctgac 826

Db 1006 GKCKYTTTTKTTGTTGTYCSCSGGYTTYCTTKGVBCKSCCGSGSVMGSGSCGCC 947

QY 827 tattccagtcgataltggtgactagtcgcggtgctcttggtggcggtgtga 882

Db 946 CGTYBGHGTCTSKKTTTCTTCTPKRYTTCYTTYGYTCTPKYHRCDTCTSSM 891

RESULT 9

CNS010C9

LOCUS

DEFINITION

CNS010C9 970 bp DNA GSS 26-JUL-1999
Drosophila melanogaster genome survey sequence SP6 end of BAC
BACN03J05 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
<http://www.edgp.ebi.ac.uk> - This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelOBAC11.

location/Qualifiers
1. .970
/organism="Drosophila melanogaster"
/plasmid="pBelOBAC11"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"


```
Db 1158 CSEKSGCCBGSKBBSYSSKSCSSCCCYCBGTGCCCCCSCSGSGKSTWCCSCYCBYKT 1099
QY 673 ctggcgatcgctgcagcagtgctgagcagctatgtctgtatcctacclacagtcgcgt 732
Db 1098 KKGKGGKGGKKBKTCYSYKKTBBGGGGKGGSTGTGTCTCYBCCCCCCCCCCCCSSSGG 1039
QY 733 ggcctgctggtttgtgcgcacatgtattttgtctggcggtattataggctcta 785
Db 1038 GSSSSSSSGGGGSCCTMTYYCTTKCCCKTGGGGRGCTKTGGGGGGGBKBTY 986

RESULT 14
BG167536 1868 bp mRNA EST 06-FEB-2001
LOCUS 602342745F1 NIH_MGC_89 Homo sapiens cDNA clone IMAGE:4452947 5',
DEFINITION mRNA sequence.
ACCESSION BG167536
VERSION BG167536
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE 1 (bases 1 to 1868)
JOURNAL NIH-MGC http://mgs.nci.nih.gov/.
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM10242 row: 1 column: 12.

FEATURES
Source Location/Qualifiers
1..1868
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4452947"
/clone_lib="NIH_MGC_89"
/tissue_type="hypernephroma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

BASE COUNT 599 a 624 c 370 g 274 t 1 others
ORIGIN

Query Match 3.0%; Score 41.8; DB 11; Length 1868;
Best Local Similarity 48.5%; Pred. No. 6.4;
Matches 115; Conservative 0; Mismatches 122; Indels 0; Gaps 0;
```

```
Db 1284 GCGTGTCCAGAGAGGGTGGTGTGTTGTGCGATCTAAGCTGCGTCCCATTTCTG 1228
RESULT 15
CNS016KD 1100 bp DNA GSS 26-JUL-1999
LOCUS CNS016KD
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC
BACN16D22 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL106855
VERSION AL106855.1 GI:5624152
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Plasmid Drosophila melanogaster
REFERENCE Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
AUTHORS Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
JOURNAL Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1100)
Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelobAC11.

FEATURES
Source Location/Qualifiers
1..1100
/organism="Drosophila melanogaster"
/plasmid="pBelobAC11"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"
/clone="BACN16D22"
/note="end : SP6"

BASE COUNT 132 a 229 c 106 g 220 t 413 others
ORIGIN

Query Match 3.0%; Score 41.6; DB 13; Length 1100;
Best Local Similarity 21.9%; Pred. No. 6.1;
Matches 119; Conservative 182; Mismatches 241; Indels 2; Gaps 2;
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 10, 2002, 21:39:13 ; Search time 47.17 Seconds
(without alignments)
6741.045 Million cell updates/sec

Title: US-09-887-038-2

Perfect score: 1404

Sequence: 1 atgactgtctgtgcaaacctc.....attcagacgaaaaaatgtag 1404

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA:*

1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	49.8	3.5	7218	1 US-08-232-463-14	Sequence 14, Appl
2	47.4	3.4	2232	1 US-08-241-766-12	Sequence 12, Appl
3	47.4	3.4	3120	1 US-08-491-146-1	Sequence 1, Appl
4	47.4	3.4	3120	1 US-08-241-766-11	Sequence 11, Appl
5	47.4	3.4	3120	1 US-08-234-011-1	Sequence 1, Appl
6	47.4	3.4	3120	2 US-08-701-062A-1	Sequence 1, Appl
7	39.6	2.8	1723	1 US-08-241-766-1	Sequence 1, Appl
8	39.6	2.8	1723	1 US-08-241-766-2	Sequence 2, Appl
9	35.8	2.5	2056	4 US-09-334-601-12	Sequence 12, Appl
10	35.2	2.5	390	3 US-08-911-853-20	Sequence 20, Appl
11	35.2	2.5	390	4 US-09-479-409-20	Sequence 20, Appl
12	35.2	2.5	710	4 US-08-998-416-706	Sequence 706, App
13	35.2	2.5	17612	3 US-08-911-853-29	Sequence 29, Appl
14	35.2	2.5	17612	4 US-09-479-409-29	Sequence 29, Appl
15	35	2.5	30001	1 US-08-125-468-1	Sequence 1, Appl
16	35	2.5	30001	2 US-08-474-933-1	Sequence 1, Appl
17	34.8	2.5	43280	2 US-08-804-227C-1	Sequence 1, Appl
18	34.4	2.5	4791	4 US-08-949-155-49	Sequence 49, Appl
19	33.6	2.4	33529	4 US-09-144-085-3	Sequence 3, Appl
20	33.4	2.4	4403765	4 US-09-103-840A-2	Sequence 2, Appl
21	33.4	2.4	4411529	4 US-09-103-840A-1	Sequence 1, Appl
22	33	2.4	326	1 US-08-700-575-3	Sequence 3, Appl
23	32.8	2.3	3410	4 US-09-020-956-110	Sequence 110, App
24	32.8	2.3	3410	4 US-09-030-607-110	Sequence 110, App
25	32.6	2.3	1088	4 US-09-077-675A-6	Sequence 6, Appl
26	32.6	2.3	1122	4 US-09-077-675A-9	Sequence 9, Appl
27	32.6	2.3	2221	4 US-09-232-191-24	Sequence 24, Appl

28	32.6	2.3	2221	4 US-09-232-200-74	Sequence 74, Appl
29	32.6	2.3	2221	4 US-09-232-197-74	Sequence 74, Appl
30	32.4	2.3	1272	3 US-09-036-987A-28	Sequence 28, Appl
31	32.4	2.3	1272	4 US-09-370-700-28	Sequence 28, Appl
32	32.4	2.3	80161	3 US-09-036-987A-1	Sequence 1, Appl
33	32.4	2.3	80161	4 US-09-370-700-1	Sequence 1, Appl
34	32.2	2.3	28958	1 US-08-258-261B-6	Sequence 6, Appl
35	32.2	2.3	28958	1 US-08-456-837-6	Sequence 6, Appl
36	32.2	2.3	28958	1 US-08-457-342-6	Sequence 6, Appl
37	32.2	2.3	28958	1 US-08-457-646A-6	Sequence 6, Appl
38	32.2	2.3	28958	1 US-08-458-076A-6	Sequence 6, Appl
39	32.2	2.3	28958	1 US-08-764-233A-4	Sequence 4, Appl
40	32.2	2.3	28958	1 US-08-457-335A-6	Sequence 6, Appl
41	32.2	2.3	28958	1 US-08-729-214-6	Sequence 6, Appl
42	32.2	2.3	28958	3 US-09-028-934-6	Sequence 6, Appl
43	32.2	2.3	49377	1 US-08-764-233A-1	Sequence 1, Appl
44	32	2.3	6343	3 US-08-581-148C-30	Sequence 30, Appl
45	31.8	2.3	1548	2 US-08-762-106-5	Sequence 5, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: ptz9pt-fls
; US-08-232-463-14


```
STRANDEDNESS: single
TOPOLOGY: both
MOLECULE TYPE: DNA
DESCRIPTION:
HYPOTHETICAL: NO
ANTI-SENSE:
FRAGMENT TYPE:
ORIGINAL SOURCE: inhA operon
ORGANISM: M tuberculosis
STRAIN:
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL TYPE:
CELL LINE:
ORGANELLE:
IMMEDIATE SOURCE: M tuberculosis
POSITION IN GENOME:
CHROMOSOME/SEGMENT:
MAP POSITION:
UNITS:
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
PUBLICATION INFORMATION: No. 5556778e
AUTHORS:
TITLE:
JOURNAL:
VOLUME:
PAGES:
DATE:
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO:
US-08-491-146-1

Query Match          3.4%; Score 47.4; DB 1; Length 3120;
Best Local Similarity 54.2%; Pred. No. 0.00062;
Matches 96; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 1070 cgttagcgccactccgctcccgctggaagtcggttagggcgagactagggcttga 1129
      ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1041 CGTCGATCGCGTTTCACATCGCACGGGTAGCCCAAGAGCAGGCGCCAGCTGGTGTCTCA 1100

QY 1130 cggccttcgctgctgctgctgctgacgagcggtgacgagcggtgacgagcggtgac 1189
      ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1101 CCGGTTTCGACCGCGCTGCTGATTCAGCGCATCACCGACCGCTGCCGCAAGGCC 1160

QY 1190 tgcgagcgatcgcaatccccaagccttggtagtgtagcttgccggttg 1246
      ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1161 CGCTGCTCGAAGTCAAGTCAAGCAAGAGAGACCACTGGCCAGCTTGGCCGCGGG 1217

RESULT 4
US-08-241-766-11
; Sequence 11, Application US/08241766
; Patent No. 5686590
; GENERAL INFORMATION:
; APPLICANT: JACOBS, W. R.
; APPLICANT: COLLINS, D. M.
; APPLICANT: BANERJEE, A.
; APPLICANT: DELISLE, G. W.
; APPLICANT: WILSON, T. M.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTING
; TITLE OF INVENTION: AND TREATING MYCOBACTERIAL INFECTIONS USING AN inhA AGENT
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
```

```
STREET: 755 Page Mill Road
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentia Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/241,766
FILING DATE: 12-MAY-1994
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: MONROY, GLADYS H.
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 25237-20003.20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 813-5600
TELEFAX: (415) 494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 3120 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-241-766-11
```

```
Query Match          3.4%; Score 47.4; DB 1; Length 3120;
Best Local Similarity 54.2%; Pred. No. 0.00062;
Matches 96; Conservative 0; Mismatches 81; Indels 0; Gaps 0;
```

```
QY 1070 cgttagcgccactccgctcccgctggaagtcggttagggcgagactagggcttga 1129
      ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1041 CGTCGATCGCGTTTCACATCGCACGGGTAGCCCAAGAGCAGGCGCCAGCTGGTGTCTCA 1100

QY 1130 cggccttcgctgctgctgctgctgacgagcggtgacgagcggtgacgagcggtgac 1189
      ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1101 CCGGTTTCGACCGCGCTGCTGATTCAGCGCATCACCGACCGCTGCCGCAAGGCC 1160

QY 1190 tgcgagcgatcgcaatccccaagccttggtagtgtagcttgccggttg 1246
      ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1161 CGCTGCTCGAAGTCAAGTCAAGCAAGAGAGACCACTGGCCAGCTTGGCCGCGGG 1217
```

```
RESULT 5
US-08-234-011-1
; Sequence 1, Application US/08234011
; Patent No. 5702935
; GENERAL INFORMATION:
; APPLICANT: Sacchetti et al
; TITLE OF INVENTION: METHOD AND COMPOUNDS FOR INHIBITING
; TITLE OF INVENTION: LIPID BIOSYNTHESIS OF BACTERIA AND
; TITLE OF INVENTION: PLANTS
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amster, Rothstein & Ebenstein
; STREET: 90 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Word Processor (ASCII)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/234,011
; FILING DATE: No. 5702935 yet assigned
```



```

: CLASSIFICATION: 435
: PRIOR APPLICATION DATA: none
: ATTORNEY/AGENT INFORMATION:
:   NAME: Pasqualini, Patricia A.
:   REGISTRATION NUMBER: 34,894
:   REFERENCE/DOCKET NUMBER: 96700/271
:   TELECOMMUNICATION INFORMATION:
:     TELEPHONE: (212) 697-5995
:     TELEFAX: (212) 286-0854 or 286-0082
:     TELEX: TWX 710-581-4766
: INFORMATION FOR SEQ ID NO: 1:
:   SEQUENCE CHARACTERISTICS:
:     LENGTH: 3120
:     TYPE: nucleic acid
:     STRANDEDNESS: single
:     TOPOLOGY: both
:     MOLECULE TYPE: DNA
:     DESCRIPTION:
:     HYPOTHETICAL: NO
:     ANTI-SENSE:
:     FRAGMENT TYPE:
:     ORIGINAL SOURCE: inhA operon
:     ORGANISM: M tuberculosis
:     STRAIN:
:     INDIVIDUAL ISOLATE:
:     DEVELOPMENTAL STAGE:
:     HAPLOTYPE:
:     TISSUE TYPE:
:     CELL TYPE:
:     CELL LINE:
:     ORGANELLE:
:     IMMEDIATE SOURCE: M tuberculosis
:     POSITION IN GENOME:
:     CHROMOSOME/SEGMENT:
:     MAP POSITION:
:     UNITS:
:     FEATURE:
:     NAME/KEY:
:     LOCATION:
:     IDENTIFICATION METHOD:
:     OTHER INFORMATION:
:     PUBLICATION INFORMATION: NO. 5702935e
:     AUTHORS:
:     TITLE:
:     JOURNAL:
:     VOLUME:
:     PAGES:
:     DATE:
:     DOCUMENT NUMBER:
:     FILING DATE:
:     PUBLICATION DATE:
:     RELEVANT RESIDUES IN SEQ ID NO:
:     US-08-234-011-1

```

```

Query Match 3.4%; Score 47.4; DB 1; Length 3120;
Best Local Similarity 54.2%; Pred. No. 0.00062;
Matches 96; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

```

```

QY 1070 cyttgagcgccctactccgtcccgctggaagtcgcyttgagggcggaactactggcttga 1129
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1041 CGTCGATCGCGCTTTCACATCGCACGGGTAGCCCAAGAGCAGGGCCGCCCAAGCTGCTCA 1100
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1130 cggccttcgcttgctgctgtgtgcaagcgcytgaagcggtgcgaggtgagccgac 1189
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1101 CCGGGTTGACCGGCTGCGGCTGATTCAGCGCATCACCGACCGGCTGCCGCAAGGCC 1160
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1190 tgcgagcgagtcgcaatccccaagccttgggtgatgtagctgagccgttgg 1246
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1161 CGCTGCTGAACTCGACGTGCAAAACGAGAGCACCTGGCCACCTGGCCGCGGG 1217

```

RESULT 6

```

US-08-701-062A-1
: Sequence 1, Application US/08701062A
: Patent No. 5882878
: GENERAL INFORMATION:
:   APPLICANT: Sacchetti et al
:   TITLE OF INVENTION: inhA CRYSTALS AND THREE
:   TITLE OF INVENTION: DIMENSIONAL STRUCTURE
:   NUMBER OF SEQUENCES: 1
:   CORRESPONDENCE ADDRESS:
:     ADDRESSEE: Amster, Rothstein & Ebenstein
:     STREET: 90 Park Avenue
:     CITY: New York
:     STATE: New York
:     COUNTRY: USA
:     ZIP: 10016
: COMPUTER READABLE FORM:
:   MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
:   COMPUTER: IBM PC Compatible
:   OPERATING SYSTEM: MS-DOS
:   SOFTWARE: Word Processor (ASCII)
:   CURRENT APPLICATION DATA:
:     APPLICATION NUMBER: US/08/701,062A
:     FILING DATE: August 21, 1996
:     CLASSIFICATION: 435
:     PRIOR APPLICATION DATA: none
:     ATTORNEY/AGENT INFORMATION:
:       NAME: George, Kenneth P
:       REGISTRATION NUMBER: 30,259
:       REFERENCE/DOCKET NUMBER: 96700/296
:       TELEPHONE: (212) 697-5995
:       TELEFAX: (212) 286-0854 or 286-0082
:       TELEX: TWX 710-581-4766
: INFORMATION FOR SEQ ID NO: 1:
:   SEQUENCE CHARACTERISTICS:
:     LENGTH: 3120
:     TYPE: nucleic acid
:     STRANDEDNESS: single
:     TOPOLOGY: both
:     MOLECULE TYPE: DNA
:     DESCRIPTION:
:     HYPOTHETICAL: NO
:     ANTI-SENSE:
:     FRAGMENT TYPE:
:     ORIGINAL SOURCE: inhA operon
:     ORGANISM: M tuberculosis
:     STRAIN:
:     INDIVIDUAL ISOLATE:
:     DEVELOPMENTAL STAGE:
:     HAPLOTYPE:
:     TISSUE TYPE:
:     CELL TYPE:
:     CELL LINE:
:     ORGANELLE:
:     IMMEDIATE SOURCE: M tuberculosis
:     POSITION IN GENOME:
:     CHROMOSOME/SEGMENT:
:     MAP POSITION:
:     UNITS:
:     FEATURE:
:     NAME/KEY:
:     LOCATION:
:     IDENTIFICATION METHOD:
:     OTHER INFORMATION:
:     PUBLICATION INFORMATION: NO. 5882878e
:     AUTHORS:
:     TITLE:
:     JOURNAL:
:     VOLUME:
:     PAGES:
:     DATE:
:     DOCUMENT NUMBER:
:     FILING DATE:

```



```

: NUMBER OF SEQUENCES: 37
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genencor International
: STREET: 925 Page Mill Road
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94304-1013
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/479,409
: FILING DATE:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/911,853
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Glaister, Debra J
: REGISTRATION NUMBER: 33,888
: REFERENCE/DOCKET NUMBER: GC361-2
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 650-846-7620
: TELEFAX: 650-845-6504
: INFORMATION FOR SEQ ID NO: 29:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 17612 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
:
: US-09-479-409-29
:
: Query Match 2.5%; Score 35.2; DB 4; Length 17612;
: Best Local Similarity 58.7%; Pred. No. 4.1;
: Matches 61; Conservative 0; Mismatches 43; Indels 0; Gaps 0;
:
: QY 1143 gctgctgtgltcacgagcgtgacggcggtgacgagcgtgacgcgactgcggcgatcg 1202
: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
: Db 14081 GATGCTGCTGTACACCGTACGGGTGGCGCTGCGCGCGGAGCGCGGCTGCAGGGCAAGAT 14140
:
: QY 1203 caatcccaagccctttgttgatgtagcttgccggtttgg 1246
: | | | | | | | | | | | | | | | | | | | | | | | | | |
: Db 14141 CGAAGACCATGCCCTGTGACCCCTGAGTGGCTTCGTGGGGGTG 14184
:
: RESULT 15
: US-08-125-468-1
: Sequence 1, Application US/08125468
: Patent No. 5589385
: GENERAL INFORMATION:
: APPLICANT: Ryan, Michael J.
: APPLICANT: Lotvin, Jason A.
: APPLICANT: Strathy, Nancy
: APPLICANT: Fantini, Susan E.
: TITLE OF INVENTION: Cloning of the biosynthetic pathway for
: TITLE OF INVENTION: chlortetracycline and tetracycline Formation and cosmids
: TITLE OF INVENTION: useful therein
: NUMBER OF SEQUENCES: 1
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: American Cyanamid Company
: STREET: One Cyanamid Plaza
: CITY: Wayne
: STATE: New Jersey
: COUNTRY: USA
: ZIP: 07470
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25

```

```

1 CURRENT APPLICATION DATA:
2 APPLICATION NUMBER: US/08/125,468
3 FILING DATE: 22-SEP-1993
4 CLASSIFICATION: 435
5 ATTORNEY/AGENT INFORMATION:
6 NAME: Tsevdos, Estelle J
7 REGISTRATION NUMBER: 31,145
8 REFERENCE/DOCKET NUMBER: 31,255-02
9 TELECOMMUNICATION INFORMATION:
10 TELEPHONE: (201)831-3241
11 TELEFAX: (201)831-3305
12 INFORMATION FOR SEQ ID NO: 1:
13 SEQUENCE CHARACTERISTICS:
14 LENGTH: 30001 base pairs
15 TYPE: nucleic acid
16 STRANDEDNESS: single
17 TOPOLOGY: linear
18 MOLECULE TYPE: DNA (genomic)
19 US-08-125-468-1

```

[illegible]

Search completed: January 10, 2002, 22:26:54
Job time: 2861 sec

1
